Top Quark Production in Multi-Jet Final States

by

Eunil Won

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Thomas Ferbel

Department of Physics and Astronomy

The College

Arts and Sciences

University of Rochester

Rochester, New York

1996

Curriculum Vitae

The author was born in Seoul, Korea, on September 20, 1969. He attended the Korea University from 1988 to 1992. He graduated from the Korea University with a Bachelor of Science degree in 1992. He came to the University of Rochester in the spring of 1993 and began graduate studies in Physics. He received a Graduate Student Fellowship from the Department of Energy and a Graduate Student Fellowship from the Korean Foundation for Advanced Studies in 1993, 1994, 1995, and 1996. He pursued his research in Particle Physics under the direction of Professor Thomas Ferbel and received the Master of Arts degree in 1995.

Acknowledgements

Most of all, I was lucky to meet Tom Ferbel, my advisor. His enthusiasm for physics always far exceeded mine. I learned a lot about high energy physics from him. Recent discussions on Bayesian statistics were invaluable and enabled me to digest new concepts as a scientist. He has also provided me excellent editorial assistance of all of my almost Korean-like English without any complaints. I thank him for his guidance on every aspect of high energy physics and for his rational and critical advice of all steps in my research.

My working experience in DØ started with Stefan Grünendahl at the ground floor of DAB. First, he taught me how CAMAC ADC works by drawing cartoons, and years later he again taught me about FASTBUS (and how frequently it fails!) at LAB6. I found his help exceptional. I thank him for his kind and continuous effort to teach me various hardware issues. I also thank Jadwiga Warchol, Alan Bross, and all other members of the fiber tracking group for helping me learn about experimental physics.

I met Chip Stewart after I started working on mass fitting issues. His knowledge of DØ software surprised me every day. He also continuously provided me new ideas on mass fitting issues and new ways to tweak PAW, etc. It was a pleasure to work with him because of his sincerity and interest in research. I thank him for his excellent guidance in the all-jets analysis. Without him, I wouldn't be able to write this thesis. I also thank to Norm Amos for his help, after he took over Chip's responsibility to lead the all-jets group.

I thank Heidi Schellman and Shuichi Kunori for guiding me in developing the monitoring processes on the DØ production farm. It was quite an unique opportunity to be involved in DØ farm-software improvement using UNIX stuff.

When Tom Ferbel was on sabbatical leave, Harrison Prosper monitored my

Abstract

We have searched for $t\bar{t}$ production in multi-jet final states in $p\bar{p}$ collisions at a center-of-momentum energy of 1.8 TeV. Each of the top quarks in these states decays predominantly to a bottom quark and a W boson, with the Ws decaying into lighter quark-antiquark pairs. Although 44 % of all $t\bar{t}$ production involves such multi-jet final states, the background from Quantum Chromodynamic (QCD) processes is an overwhelming factor of 1000 larger, making it difficult to extract evidence for a signal.

Our study was performed using the DØ detector at the Tevatron $p\bar{p}$ collider at Fermilab, with a data set corresponding to 95.3 events/pb. After imposing selection criteria to enhance the signal relative to background, three different analysis techniques provided evidence for a $t\bar{t}$ signal. However, the observed excess is not significant enough to establish the unambiguous presence of a signal in the multi-jet channel. However, the yield of 7.9 ± 7.1 pb (at $m_t = 160$ GeV/c^2) is consistent with expectation based on other $t\bar{t}$ channels. Assuming that the observed excess of events is due to $t\bar{t}$ production, we also attempted to extract the mass of the top quark. Due to similarities in the shape of mass spectra for background and signal, and the low signal to background ratio, we were only able to extract a value of the top mass with a large uncertainty. We obtain 148 ± 52 GeV/c^2 as our best estimate of the mass of the top quark, a value that is consistent with measurements in cleaner channels.

Contents

List of Tables

List of Figures

Foreword

The DØ experiment is a large collaborative effort, in which more than 450 physicists and students from 48 institutions are currently participating. The University of Rochester group is active in monitoring of liquid-argon purity, in data acquisition and reconstruction efforts, and in analysis of data in the areas of QCD, top-quark physics, and new phenomena. The group is also participating in development and construction projects for a scintillating-fiber tracking chamber to be included in the upgrade of the DØ detector.

I have been involved in work in several of these areas. I participated in early research and development projects for the scintillating fiber tracker by developing quality control procedures for photodetector chips, helping to develop techniques for the construction of ribbons of scintillating fibers, by performing studies of photodetector characteristics, by participating in studies of optical connectors, and by participating in studies of temperature controllers for the photodetector test. In addition, I participated in setting up electronics for the cosmic-ray test of a prototype scintillating fiber tracker. During the 1994–96 running period of the Tevatron, I also participated in operating the offline processor farm that reconstructed events from the detector signals, and have had primary responsibility for maintaining and improving the control processes and monitoring processes that handle communication between the server and worker nodes on the farm. My analysis project has been studies of top quark production in multi-jet final states.

Chapter 1

Introduction

This dissertation is devoted to an experimental search for the top quark and a measurement of its mass in multi-jet final states of proton-antiproton collisions.

All phenomena in our world appear to be explainable by four basic forces: strong, electromagnetic, weak and gravitational, and great progress has been made in elementary particle physics over the past three decades, both experimentally and theoretically, in unifying the first three forces. The "Standard Model" is the mathematical formulation of the current status of particle interactions, and in this chapter we will describe it briefly, with emphasis on the physics of the top quark in the context of the Standard Model.

1.1   Standard Model

The Standard Model is a gauge field theory based on vector fields associated with the group $SU(3) \otimes SU(2) \otimes U(1)$, on fermions representing quarks and leptons, and on at least one scalar Higgs multiplet that is needed to give vector bosons and fermions their finite masses[1, 2]. The simplest possibility has a single Higgs

doublet, and is frequently referred to as the *minimal* Standard Model[3]. This is what we will refer to as the Standard Model. The $SU(3)$ gauge interactions are the strong "color" forces associated with quantum chromodynamics (QCD), while the $SU(2) \otimes U(1)$ interactions describe the electroweak isospin and hypercharge of the Glashow-Weinberg-Salam Model[4]. Electroweak interactions are described by an $SU(2)$ weak field W^a_μ and a $U(1)$ hypercharge field B_μ, together with a complex Higgs field ϕ that transforms as a doublet under $SU(2)$. Neglecting fermions and color forces, the electroweak Lagrangian can be written as:

$$\mathcal{L} = -\frac{1}{4}B^2_{\mu\nu} - \frac{1}{4}W^{a2}_{\mu\nu} + D_\mu\phi^\dagger D^\mu\phi - V(\phi), \tag{1.1}$$

where

$$B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.2}$$

$$W^a_{\mu\nu} \equiv \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2\epsilon_{abc}W^b_\mu W^c_\nu, \tag{1.3}$$

g_2 is a (arbitrary) strength of the coupling to the weak isospin current, ϵ_{abc} is the structure constant of the group $SU(2)$, and ∂_μ is the four-gradient defined as $\partial_\mu \equiv (\partial/\partial t, \nabla)$. Quite generally, the covariant derivative is defined as

$$D^\mu \equiv \partial_\mu - \frac{1}{2}ig_1 B_\mu Y - ig_2 W^a_\mu T^a, \tag{1.4}$$

where Y is the hypercharge generator, g_1 is a (arbitrary) strength of the coupling to the weak hypercharge current, and T^a is the $SU(2)$ generator appropriate to the field on which the derivative is acting.

For the Higgs doublet, the hypercharge Y_ϕ is $+1$ and $T^a = 1/2 \, \sigma^a$, where σ^a are the standard Pauli matrices. The potential for the Higgs field (ϕ) can be written

as:

$$V(\phi) = \lambda(\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi, \tag{1.5}$$

where λ is the self-coupling and μ is a mass parameter. The mass term enters with the "wrong" sign, triggering "spontaneous symmetry breaking"[5] so that in the ground state,

$$\phi^\dagger \phi = \frac{1}{2} v^2, \tag{1.6}$$

where in tree level approximation $v^2 = \mu^2/\lambda$. Now, in a gauge where $\phi = \sqrt{1/2}(0,v)^T$, the mass matrix for the vector bosons is, to first approximation, associated with the following quadratic terms in the Lagrangian[6],

$$\frac{1}{4}(g_1 B_\mu + g_2 W^a_\mu \sigma^a)\phi^\dagger (g_1 B^\mu + g_2 W^{b\mu}\sigma^b)\phi. \tag{1.7}$$

In terms of the weak mixing angle, defined by $\tan\theta_w \equiv g_1/g_2$, the above Lagrangian can be expressed in terms of rotated fields[6]:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} \equiv \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix},$$

$$W_\mu \equiv \sqrt{1/2}(W^1_\mu + iW^2_\mu). \tag{1.8}$$

In the rotation, electromagnetic field A_μ remains massless, and couples with strength

$$e \equiv \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} = g_1 \cos\theta_w. \tag{1.9}$$

Using the rotated fields, the preceding expression (1.7) becomes[6]:

$$M_W^2 W_\mu^\dagger W^\mu + \frac{1}{2} M_Z^2 Z_\mu^2, \tag{1.10}$$

where $M_W^2 \equiv \frac{1}{4}g_2^2 v^2$ and $M_Z^2 \equiv M_W^2 / \cos^2 \theta_W$. (The relationship between the mixing angle and the masses of the physical vector boson is not true beyond tree level approximation[6].) The mass of the W boson is known to an accuracy of better than 0.3% from observations of leptonic decays of Ws produced in $\overline{p}p$ collisions at CERN[7] and Fermilab[8]. M_Z has been determined to an accuracy better than 0.1% from detailed fits to the Z^o resonance peak in e^-e^+ annihilation at SLC[9] and LEP[10]. The existence of the vector bosons, and their observed mass ratio, strongly supports the Standard Model. The mass of the Higgs boson is unknown and is a free parameter in the theory.

The fundamental fermions, the leptons and the quarks, are grouped into three "families" or "generations". The three generations of leptons, associated with the electron, the muon, and the tau, are:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad e_R; \qquad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \mu_R; \qquad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \quad \tau_R,$$

where the subscripts L and R stand for left and right-chiral objects, respectively. All neutrinos are massless in the Standard Model, and consequently no right-handed ν fields are required[6]. The right-handed leptons are therefore singlets of weak isospin. Assuming standard decays, measurements of the width of the Z^o[9, 10] exclude the existence of other than the three known generations of massless neutrinos.

Similarly, the three generations of quarks in their weak eigenstates of the unbroken gauge theory can be represented as

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \quad u_R, d_R; \qquad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad c_R, s_R; \qquad \begin{pmatrix} t \\ b \end{pmatrix}_L, \quad t_R, b_R.$$

Here, in order to assign quark mass terms for both "up-type" and "down-type" quarks, one needs not only the doublet Higgs ϕ with $Y = 1$, but also the conjugate multiplet $\tilde{\phi} = i\sigma_2\phi^*$, which transforms as a doublet with $Y = -1$. Having this possibility in the Lagrangian, provides a mixing among the quarks, with their charged weak currents given by:

$$J_{L\mu}^+ = \begin{pmatrix} u & c & t \end{pmatrix}_L \gamma_\mu V \begin{pmatrix} d \\ s \\ b \end{pmatrix} L .$$

This mixing matrix V, known as Cabibbo-Kobayashi-Maskawa matrix[11] (CKM) describes the relative strengths of the mixing of quarks in electroweak interactions.

Unlike the leptons, the quarks interact strongly, and carry a color index. All the quark fields transform as fundamental triplets under color $SU(3)$. The gluon field G_μ^a, which mediates this strong interaction, adds to the Lagrangian $\mathcal{L}$ a term

$$-\frac{1}{4}G_{\mu\nu}^{a2} \tag{1.11}$$

where $G_{\mu\nu}^a \equiv \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_3 f_{abc}G_\mu^b G_\nu^c$, with g_3 being the QCD coupling constant and f_{abc} the $SU(3)$ structure constants. Measurement of hadron production in e^+e^- interactions, and in particular the value of R, defined as $R \equiv \sigma(e^-e^+ \to hadrons)/\sigma(e^+e^- \to \mu^-\mu^+)$, confirmed that there are indeed three colors for quarks[11].

1.2   Top Quark Production and Decay

The production cross section of top-antitop pairs in hadron collisions can be written in the Born approximation as follows[12]:

$$\sigma(t\bar{t}) = \sum_{i,j=q\bar{q},\bar{q}q,gg} \int_0^1 dx_1 dx_2 F^{(1,i)}(x_1)F^{(2,j)}(x_2)\hat{\sigma}^{(ij)}\left(\frac{4m^2}{x_1 x_2 s}\right), \qquad (1.12)$$

where m is the mass of the top quark, and s is the square of the total center-of-momentum energy. The function $F^{(i,k)}$ are the parton (partons refer to both quarks and gluons) densities for parton k in hadron i, and are evaluated at a scale μ of the order of the heavy quark mass m. At this lowest order, the two partonic subprocesses are quark-antiquark annihilation and gluon-gluon fusion as shown in Fig. 1.1. These subprocesses are of order α_s^2 in the strong coupling strength. The calculated top production cross section depends on these subprocess cross sections $(\hat{\sigma}^{(ij)})$ and on the parton densities that specify the probability of having quarks, antiquarks, and gluons, of any specified momentum fraction x, in the incident proton and antiproton.

Both the order α_s^2, and the next-to-leading order α_s^3, contributions have been calculated by several groups[13, 14, 15]. One interesting result is that the size of the $O(\alpha_s^3)$ terms is particularly large near the $t\bar{t}$ production threshold. This, of course, raises questions about the reliability of the perturbative calculation. Because of the large mass of the top quark, this region of phase space is important for top quark production at the Tevatron. These effects have been included in the calculation of the $t\bar{t}$ total cross section using the so called resummation technique, that resums leading log terms to provide a better estimate of the threshold effect. The resulting cross section can be fitted adequately in the range of $m_t = 160$ to 190 GeV/c^2 by the expression[15]:

$$\sigma(t\bar{t}) = e^{\frac{175-m_t}{31.5}}\left(4.75^{+0.63}_{-0.68}\right) \text{ pb.} \qquad (1.13)$$
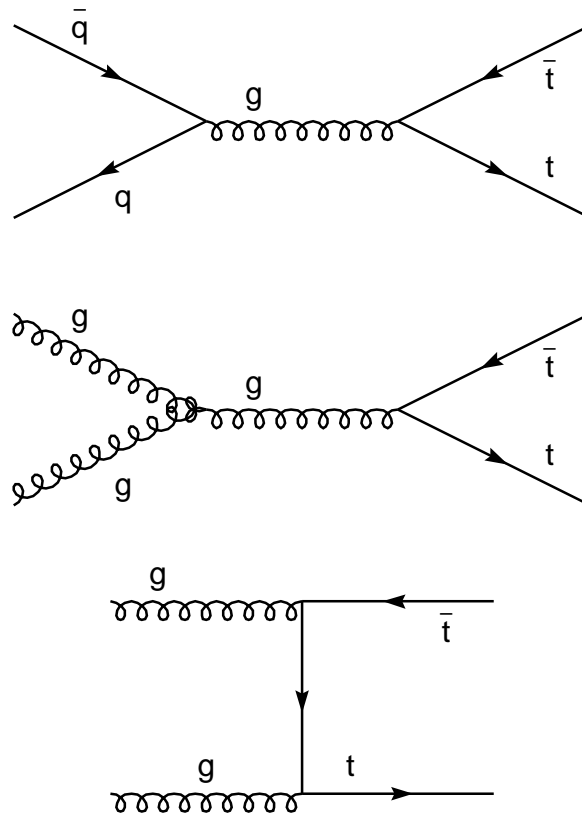
Figure 1.1: Leading order processes for $t\bar{t}$ production.

Table 1.1: Branching fractions for $t\bar{t}$ decay channels.

		$W^+ \to$			
		$q\bar{q}'$ (6/9)	$e^+\nu_e$ (1/9)	$\mu^+\nu_\mu$ (1/9)	$\tau^+\nu_\tau$ (1/9)
$W^- \to q'\bar{q}$	(6/9)	36/81	6/81	6/81	6/81
$e^-\bar{\nu}_e$	(1/9)	6/81	1/81	1/81	1/81
$\mu^-\bar{\nu}_\mu$	(1/9)	6/81	1/81	1/81	1/81
$\tau^-\bar{\nu}_\tau$	(1/9)	6/81	1/81	1/81	1/81

Figure 1.2 shows the predicted $t\bar{t}$ total cross section as a function of mass of the top quark, where the dashed curves represent the theoretical uncertainty due to unknown higher-order effects and imprecise knowledge of the physical parameters such as α_s.

In the Standard Model, the top quark decays almost exclusively to an on-shell W boson and a b quark (for $m_t > m_W + m_b$)[17, 18]. The W can decay either semileptonically, for example, $W \to e + \nu$, or hadronically, $W \to \bar{q} + q'$, while the b quark forms a "jet". Due to the fact that the mass of the W is much greater than that of its decay products, to a good approximation, each allowed decay mode of the W boson is equally probable, except that the hadronic modes are three times more likely because of the color factor. This results in the predictions of Table 1.1 for different $t\bar{t}$ branching fractions.

In this dissertation, we will focus on the extraction of the signal for the top
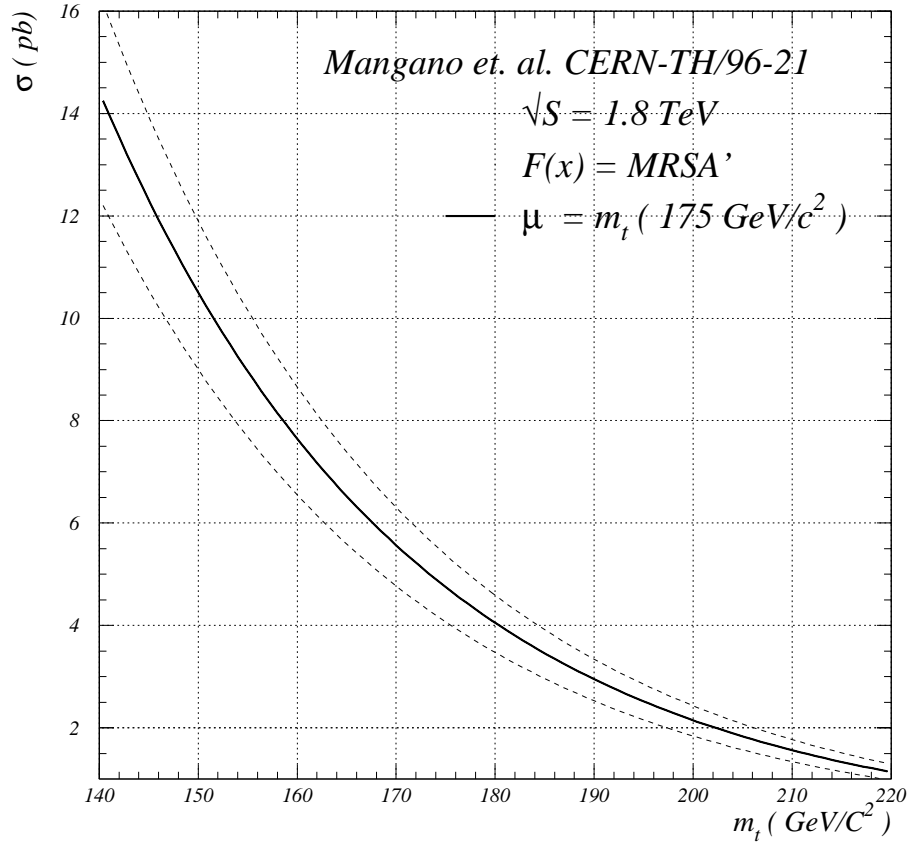
Figure 1.2: The top-antitop production cross section as a function of top mass for $\sqrt{(s)}$ = 1.8 TeV[15]. The solid line shows the central value for the prediction, and the dashed lines show the theoretical uncertainty. The Martin-Roberts-Stirling set A' (MRSA')[16] parton distribution was used in the calculation.

quark, and the reconstruction of its mass, in states where the W bosons from top and antitop quarks decay only into hadronic jets. This channel, the "all-jets" final state, has an advantage of having the largest branching fraction (4/9). Apart from the largest branching ratio, the measurement of the mass of the top quark in the six-jet topology has another important advantage over measuring the mass in leptonic decays. In leptonic events, one neutrino for every semi-leptonic decay escapes the detector without being measured directly. This complicates the procedure of extracting the mass in this channel and can degrade the measurement. However, the background to $t\bar{t}$ production in multi-jet events with 6 or more jets in the final state, which has its origin in the quantum chromodynamic (QCD) production of partonic jets, has a cross section of the order of 100 nb[19, 20]. The ratio of top quark events to the QCD background is approximately 5 pb / 100 nb $\approx$ 1 / 20,000, and it is therefore quite challenging to extract the top-quark signal, to measure the $t\bar{t}$ total cross section, and to measure the mass of the top quark in this channel. One difference between top quark events and the QCD background, however, is that the b-quark and quarks from W decay are produced with higher transverse momentum (p_T) values than partons in typical QCD processes. Figure 1.3 (a) shows the transverse energy (E_T) distributions of b-quarks and light quarks from W decay in top quark production at the Tevatron. (We will usually assume that the mass of the light quarks can be neglected, and that the mass of the b-quark is 5 GeV/c^2, and define the transverse energy E_T as $\sqrt{p_T^2 + m_q^2}$.) For comparison, the E_T distribution for partons from background QCD processes is also shown in the plot. On the average, quarks from top quark decay have higher E_T values. Also, the pseudorapidity (η) distributions of quarks from top quark decay and partons from QCD background are significantly different, as is shown in Fig. 1.3 (b). Such differences will be utilized in our analysis.

In addition to quarks expected from the leading-order processes in $t\bar{t}$ produc-
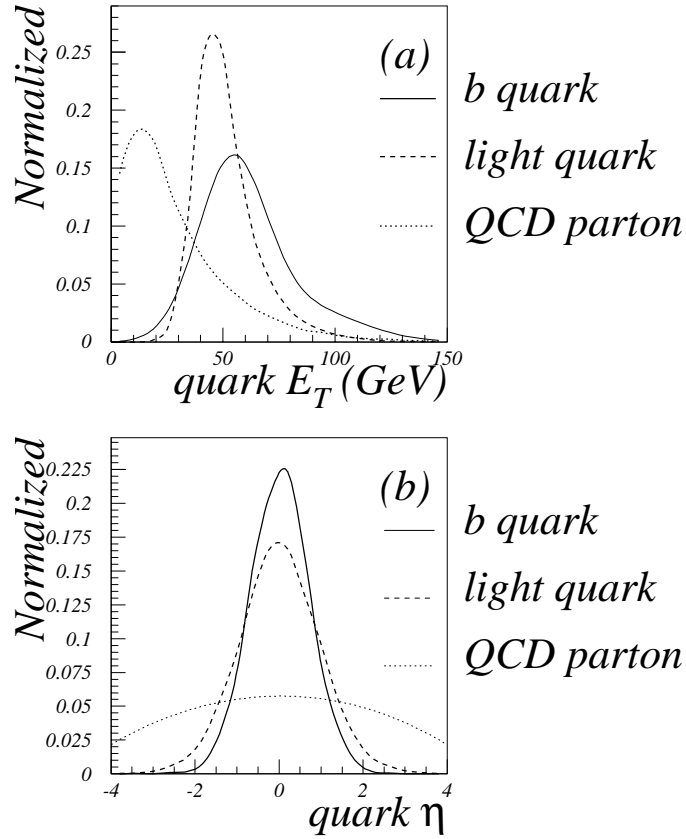
Figure 1.3: Kinematic properties of quarks from top-quark decay and from standard QCD process. For the calculation of E_T and η of the partons, we used the HERWIG program. Solid lines are for the b quark, and dashed lines for quarks from W decay in $t\bar{t}$ events. The dotted lines are for QCD background processes. (a) E_T distributions of partons, (b) η distributions of partons.
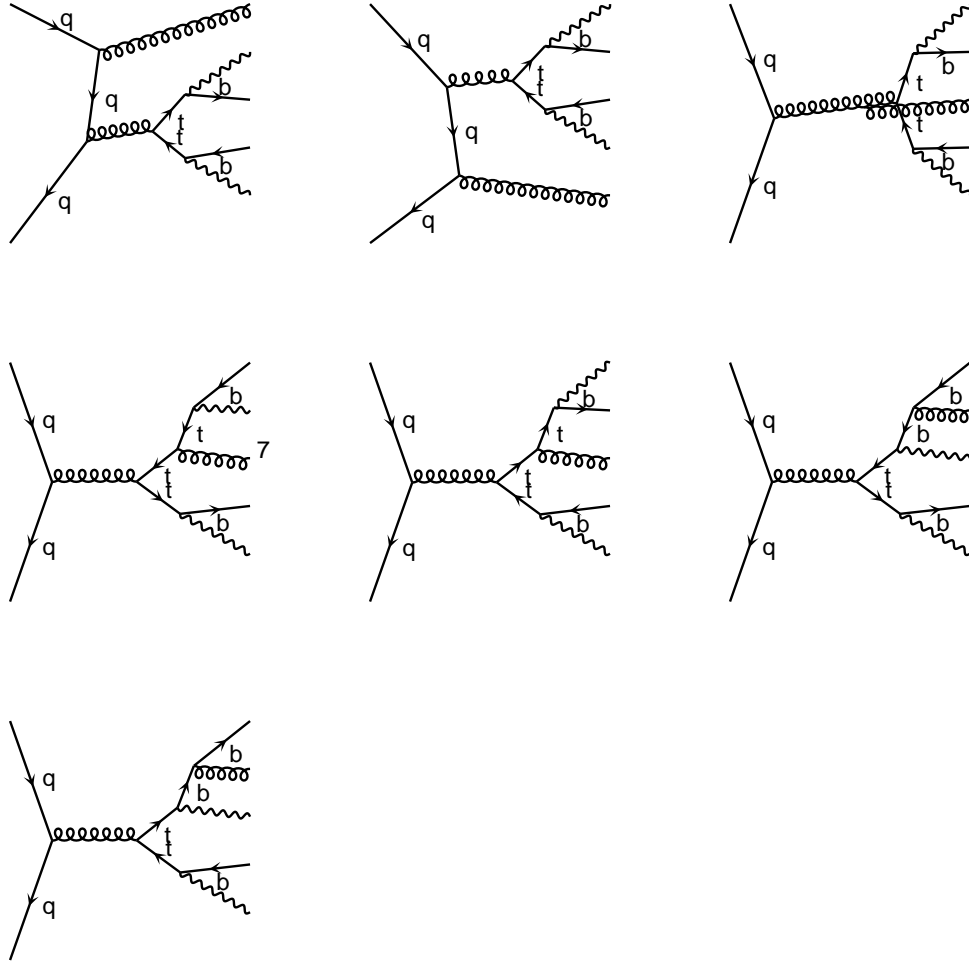
Figure 1.4: Seven Feynman diagrams that dominate $t\bar{t}$ production involving extra gluons[21].

tion, extra partons can be produced, for example, as gluons emitted from the incoming partons, from the top quarks before they decay, or from the b quarks in the final state. There have been several studies of the production of extra jets in $t\bar{t}$ final states, and there is now available a complete calculation of the exact matrix element for the process, including contributions from gluons emitted during production and decay and the interference between these at the tree-level[21]. Figure 1.4 shows seven diagrams that dominate $q\bar{q} \rightarrow t\bar{t}$ production at Tevatron energies. (Because of gluon density suppression in the proton at high x, $q\bar{q}$ annihilation cross section is almost an order of magnitude larger than that for gg fusion for $m_t \approx 175$ GeV/c^2.) The presence of extra gluons affects not only the determination of the mass of the top quark by worsening the resolution on the Wb invariant mass, but also by increasing the combinatoric backgrounds in selecting the correct set of jets to form the top mass. This issue will be addressed later in this dissertation.

1.3   Hadronization

Colored quarks and gluons can be regarded as essentially free during a hard collision, but color forces subsequently force them into colorless hadrons. This is called *hadronization* or *fragmentation*[22, 23]. This fragmentation of partons into bundles of hadrons that form jets is a non-perturbative phenomena and cannot at present be calculated from first principle. The process must therefore be described phenomenologically, guided by general principles and physical ideas.

A simple mathematical model due to R.P. Feynman and R.D. Field[24] can be used to parametrize the non-perturbative aspects of quark and gluon jets. This so-called Feynman Field fragmentation model, assumes that jets can be analyzed on the basis of a recursive principle. The ansatz is based on the idea that the

fragmenting quark is combined with an antiquark, originating from a $q\overline{q}$ pair created out of the vacuum, to give a "first-generation" meson with a parton energy fraction z. The leftover quark, with energy fraction $(1-z)$, is fragmented in the same way, and so on, until the leftover energy falls below some arbitrary cutoff. For gluon fragmentation, the gluon is first split into a quark-antiquark pair. Then by assigning with equal probability all of the gluon's momentum to one or the other quark, the gluon behaves in effect as a quark of random flavor. This is the basic way that fragmentation occurs in the event generator, ISAJET[25].

However, when a color-neutral $q\overline{q}$ pair is produced, a color-force field is created between the quarks. It fact, for a confining theory such as QCD, the color lines of force should be mostly concentrated in a narrow tube connecting the q and $\overline{q}$, thereby acting like a string with constant tension. This is the picture used in another event generator, HERWIG (Hadron Emission Reactions With Interfering Gluons)[26].

These two different models for fragmentation in top quark production will be used in our analysis of the data. Our results often depend on details of those models, and, in some cases, the two models can provide significantly different results. Such differences will limit the systematic uncertainty of our measurements.

1.4   Outline of Dissertation

In Chapter 2, we describe the Tevatron collider and the DØ detector. Particle identification is discussed in Chapter 3, and selection of the data samples used in our search in Chapter 4. Chapter 5 is devoted to the techniques used to distinguish a signal for $t\overline{t}$ production in the all-jets channel from backgrounds. In Chapter 6, techniques used to extract the mass of the top quark are presented. Conclusions are drawn in Chapter 7.

Chapter 2

The Experiment

2.1 The Accelerator Facilities

The Fermi National Accelerator Laboratory produces currently the world's highest-energy proton-antiproton colliding beams. The accelerator, which provides the protons and antiprotons, consists of a series of seven components: a preaccelerator, a linear accelerator, a rapid-cycling synchrotron, a debuncher, an antiproton source, the Main Ring, and the Tevatron. Figure 2.1 shows these components, and the following sections give brief descriptions of some of their functions[27].

2.1.1 The Preaccelerator

The Preaccelerator is where the eventual proton beam starts out as a pulsed 10 keV mA negative hydrogen-ion beam. Hydrogen gas is injected into a magnetron to a pressure of several hundred millitorr and energized with a few hundred volts[28]. Hydrogen gas is ionized inside the magnetron and then negative ions are created by positive ions striking the cathode. Figure 2.2 shows a schematic view of the magnetron source. After formation, some of the negative hydrogen ions
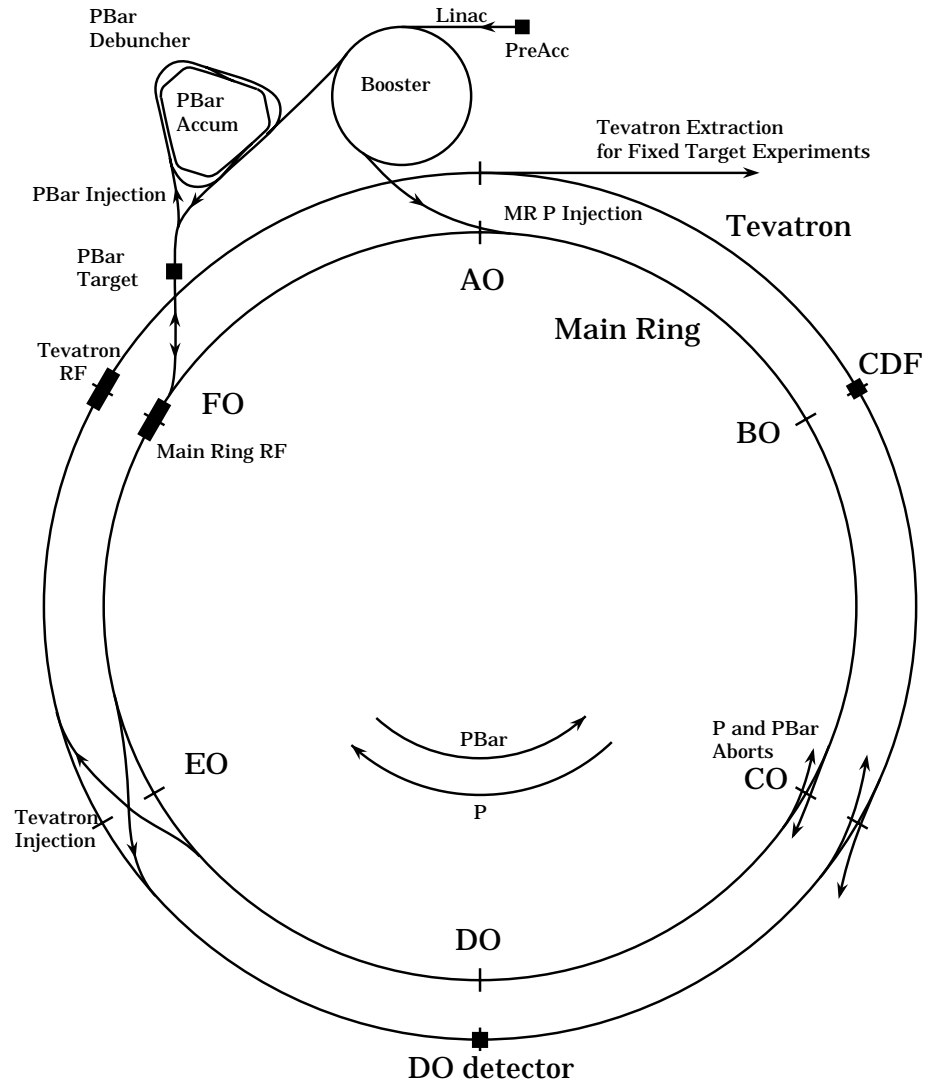
Figure 2.1: Layout of the collider facility at Fermilab (not to scale).
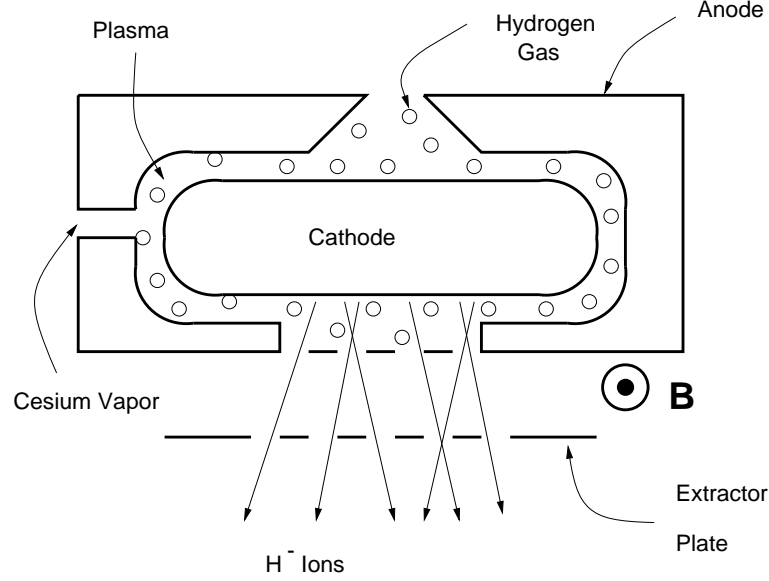
Figure 2.2: The basic configuration of a magnetron source. Cesium is introduced into the source to increase the negative ion production by lowering the surface work function[28].

are extracted through the anode aperature, are accelerated through the extraction plate to 750 keV, and transported to the Linac.

2.1.2   The Linac

The Linac is a two-stage linear accelerator that produces a pulsed beam of 400 MeV negative hydrogen ions for injection into the Booster. The first stage of the Linac, a drift-tube accelerator, accelerates the ions to 116 MeV. A new side-coupled linac has replaced a portion of the drift-tube linac and currently accelerates the beam to an energy of 400 MeV.

2.1.3   The Booster

The Booster[29] is an 8 GeV fast-cycling proton synchrotron, which serves as an injector for the Main Ring. It accelerates 400 MeV protons obtained from

ORBMP1   ORBMP2

Closed
Orbit →

Closed →

H^o To Beam Dump →

H^- →

Stripping Foil

Debunched 200 MeV
H^- Beam from Linac
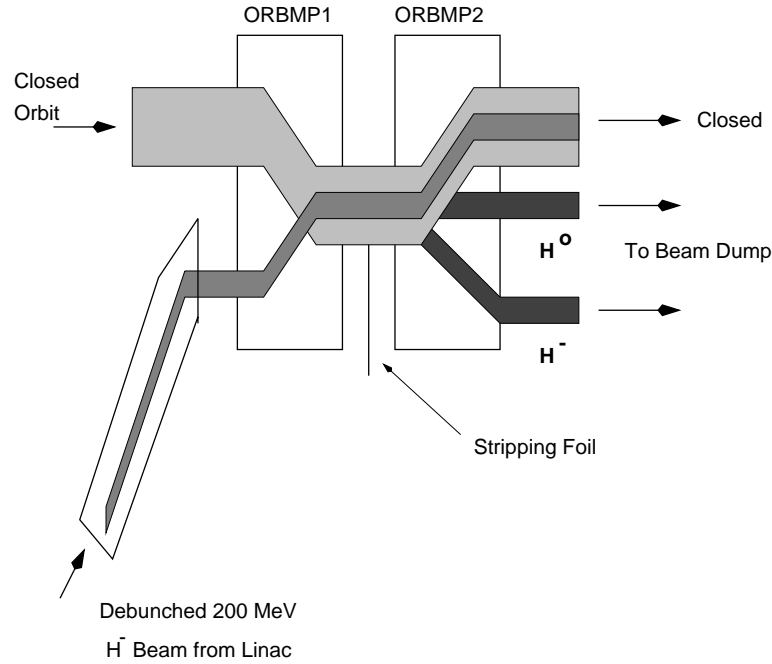
Figure 2.3: A simplified view of the Fermilab Booster injection for H^- multiturn charge-exchange injection (ORBMP refers to ORbital BuMP magnet).

the Linac via multiturn charge-exchange injection at a rate of 15 Hz. Negative hydrogen ions are brought into a parallel path with a closed orbit for protons in a straight section of the booster. The two beams are merged by passing both through two adjacent dipole magnets of opposite polarity, and then passes through a carbon foil, which strips electrons from the negative hydrogen ions, as shown schematically in Fig. 2.3. The booster operates in two modes for colliding-beam operations. When the Booster is accelerating protons for eventual injection into the Tevatron, only 11, 13, or 15 bunches are injected into the Main Ring for final coalescing into one bunch. The remaining buckets are directed to a beam dump. While antiprotons are being collected, the Booster delivers one full turn of protons to the Main Ring approximately every 2.4 sec.

2.1.4 The Main Ring

The Main Ring[30] is a 400 GeV proton synchrotron with a radius of 1000 m. Since the commissioning of the modified Tevatron in July, 1993, the Main Ring has served as a 150 GeV injector of protons and antiprotons for the Tevatron, as well as a source of 120 GeV protons used for producing antiprotons. The layout of the Main Ring was given in the Fig. 2.1, where the labels denote sections used for injection and extraction lines, and locations of colliding-beam experiments. Two major deviations from a circular orbit in the Main Ring are the vertical excursions out of the plane of the circle at BØ and DØ. The design of the Tevatron was restricted by the requirement that it should be installed within the existing Main-Ring tunnel. The exceptions to this are the two overpasses of the Main Ring at the BØ and DØ interaction regions, where the Main-Ring beamline is separated vertically from that of the Tevatron ring. The overpasses were envisioned to allow collider experiments located in these two regions to operate without interference with the detectors. The overpass built at BØ does, in fact, bypass the CDF detector at a vertical separation of approximately 19 feet. However, the vertical overpass at DØ passes through the forward muon chambers and the outer calorimeter modules.

The beam is accelerated to the Tevatron injection energy of 150 GeV, the Main Ring and the Tevatron RF systems are synchronized, and then finally the beams are injected to the Tevatron.

2.1.5 The Antiproton Source

The Antiproton Source[27] is comprised of a target station, a Debuncher ring and an Accumulator ring, and the transport lines associated with these devices. The accumulation of antiprotons involves extracting protons from the Main Ring, directing them onto a target, collecting negative-charged secondary particles, and

cooling the antiprotons so that they can be stored in the Tevatron. (Cooling means reducing the phase space occupied by the beam, and thereby fitting the beam into the smaller aperture of the Accumulator, with an accompanying reduction of the spread in the momentum) Finally, the Main-Ring RF system is synchronized to the accumulator RF system (53 MHz) and the synchronous transfer of antiprotons into the Main Ring occurs.

2.1.6   The Tevatron

The Tevatron is a proton-antiproton colliding beam synchrotron accelerator, operating at a center-of-momentum energy of 1.8 TeV. It is currently the highest energy collider in existence. All of the dipoles, quadrupoles, and correction magnets making up the basic lattice of the machine are superconducting and are cooled by liquid helium to a temperature 4.6 K. At the beginning of each store, proton bunches from the Main Ring are injected individually into the Tevatron followed by antiproton bunches. Once injected, they are ramped together to the current operating energy of 900 GeV per beam. Special superconducting quadrapoles, located on either side of the two luminous regions, squeeze the beams. This decreases the size of the beam spot to $\sigma_{x,y} \approx 40 \ \mu m$, which increases the luminosity (this is the reason that BØ and DØ are referred to as luminous regions). Table 2.1 lists several parameters of the Tevatron Run in 1992. More detailed information can be found elsewhere[27].

2.2   The Detector

The DØ detector[33] was constructed to study proton-antiproton collisions in the Fermilab Tevatron collider. The prime physics goals are the study of high-

Table 2.1: Tevatron parameters.

Accelerator radius	1000 m
Maximum beam energy	900 GeV
Injection energy	150 GeV
Peak luminosity	$\approx 10 \times 10^{30}$ cm^{-2} s^{-1}
Number of bunches	6 p, 6 $\bar{p}$
Intensity per bunch	$\approx 100 \times 10^9 p$, $\approx 50 \times 10^9 \bar{p}$
Crossing angle	0°
Bunch length	50 cm
Transverse beam radius	43 μm
Fractional Energy spread	0.15 $\times 10^{-3}$
RF frequency	53 MHz
$\bar{p}$ stacking rate	$\approx 3.5 \times 10^{10}$/hour
$\bar{p}p$ crossing frequency	290 kHz
Period between $\bar{p}p$ crossings	3.5 μs

p_T phenomena, which includes top-quark physics, heavy-boson physics, perturbative QCD, b-quark production and any unexpected phenomena. The DØ detector consists of three major parts: the tracking system, the calorimeter system, and the muon system. An isometric view of the DØ detector shows these major components in Fig. 2.4. A supporting platform (not shown) serves as the transporter for the detector to and from the DØ interaction region. Much of the front-end electronics also rests on the platform. The elevation view of the detector is given in Fig. 2.5, and shows the detector system and, in addition, the supporting platform that contains the electronics, cable connections, and service modules for power, gas and cryogens. The Tevatron beam pipe is centered on the DØ detector, while the Main Ring passes through the forward muon chambers and the outer calorimetry, as shown.

We adopt a right-handed coordinate system, in which the positive z-direction is along the incident proton beam and the y-axis is upward. The angles ϕ and θ are respectively the azimuthal and polar angles ($\theta = 0^o$ along the proton beam

Muon Chambers

Calorimeters

Tracking Chambers
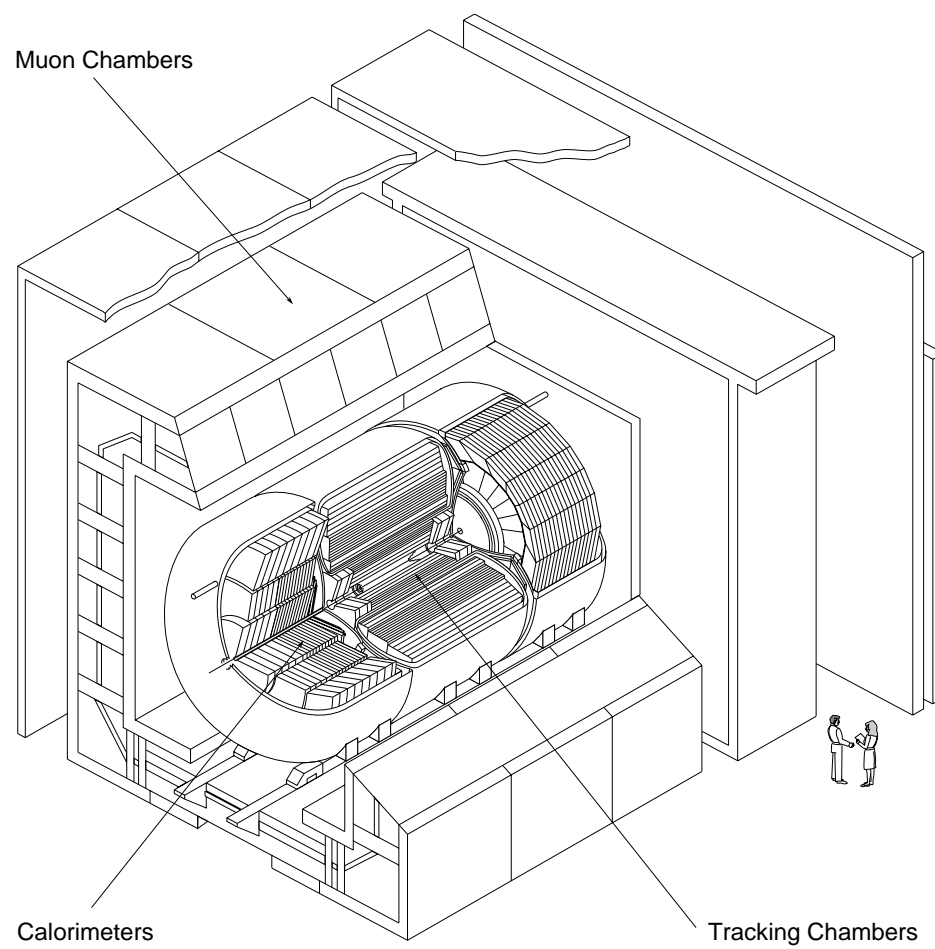
Figure 2.4: Isometric view of the DØ detector.
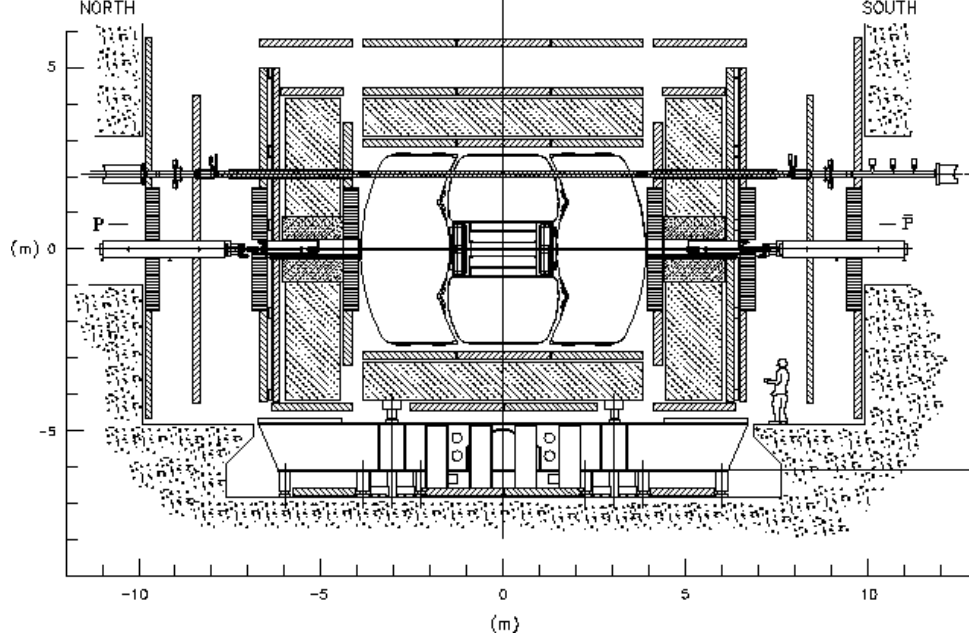
Figure 2.5: Elevation view of the DØ detector. The scales are in meters.

direction). The r-coordinate denotes the perpendicular distance from the beam axis. The pseudo-rapidity, $\eta = -\log\left(\tan\left(\theta/2\right)\right)$, approximates the true rapidity $y = 1/2\log\left((E + p_z)/(E - p_z)\right)$ in the limit that $(m/E) \to 0$, which is a useful coordinate because y (and η only approximately) is additive under Lorentz transformations along the collision axis.

2.3   Central Detectors

The DØ central detector (CD) system is composed of four subsystems: The vertex drift chamber (VTX), the transition radiation detector (TRD), the central drift chamber (CDC), and two forward drift chambers (FDC). As shown in Fig. 2.6, the VTX, TRD, and CDC are effectively three concentric cylinders that are coaxial to the beam pipe. The FDCs are oriented perpendicular to the beams so as to cover the forward region. The CD detectors extend to $r = 78$ cm and $z = \pm 135$
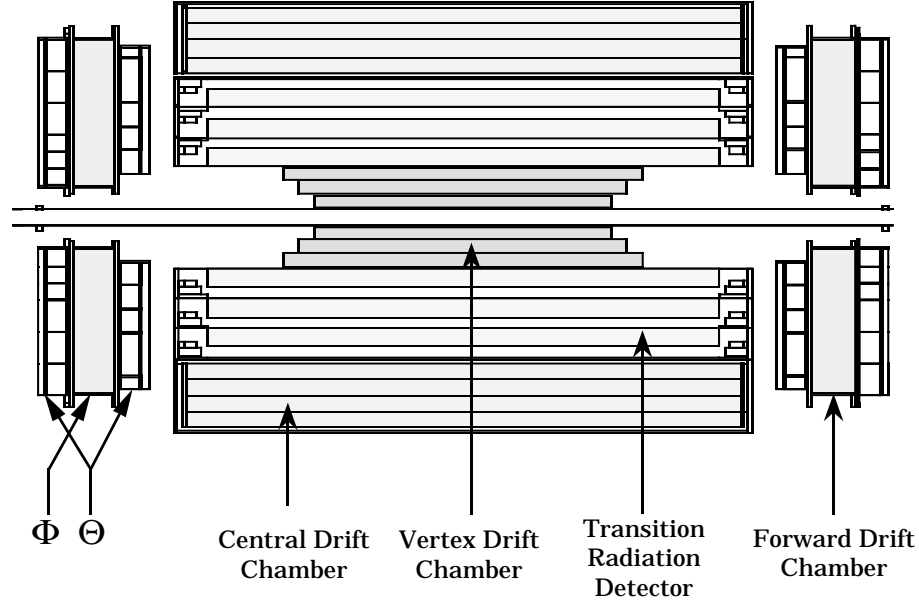
Figure 2.6: Arrangement of the DØ central detectors.

cm.

The prime considerations in the design of the DØ tracking detectors[34] were to have good two-track resolving power and high efficiency, rather than measuring momenta of charged particles. Thus, the detector has no central magnetic field. The TRD was included to help distinguish electrons from pions.

2.3.1 Vertex Drift Chamber

The passage of ionizing radiation through the gas in a cell of a chamber leaves a trail of electrons and positive ions. A drift chamber[35, 36] is a tracking device that uses the drift time of ionization-electrons in a gas to measure the spatial position of a particle that produced the ionization. Knowing the drift velocity of the electrons for a specifically chosen gas mixture, allows one to locate the position of the charged particle in the chamber, typically to an spatial accuracy of ≈ 100 μm.
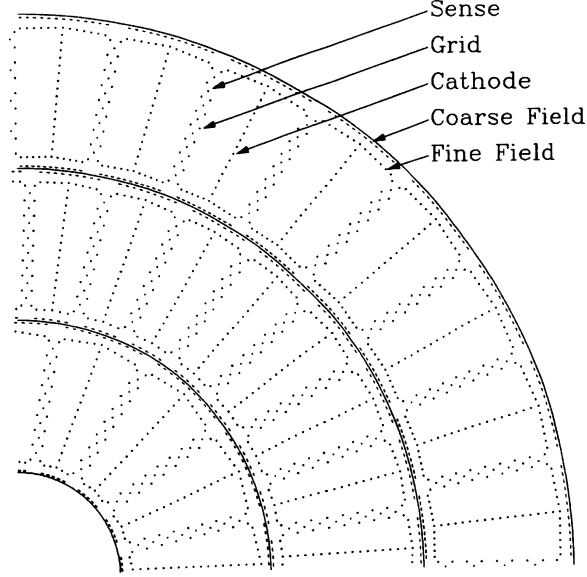
Figure 2.7: r-ϕ view of a quadrant of the VTX chamber showing the arrangement of the sense wires, grid wires and cathode field shaping electrodes.

The VTX chamber is the innermost tracking detector in DØ, with an inner radius (r) of 3.7 cm, and an outer active radius of $r = 16.2$ cm. There are three concentric layers of cells in the VTX chamber: innermost layer (VTX0, 16 cells in azimuth) and two outer layers (VTX1 and VTX2, 32 cells in azimuth). Wire sizes, composition, and other parameters of the VTX can be found in Table 2.2. As indicated in Fig. 2.7, the cells of the three layers are not aligned along the r direction. This is done to aid pattern recognition and to facilitate calibration. The sense wires also provide a measure of the z-coordinate, from the amplitudes of the signals at both ends (a method referred to as charge division)[37]. Overall, the VTX has a resolution in $r - \phi$ of 60 μm and in z of 1.5 cm[38].

2.3.2 Transition Radiation Detector

Transition radiation[39] is a type of radiation emitted when a charged particle passes between media of different dielectric or magnetic properties. (The moving

Table 2.2: VTX chamber parameters.

Length of active volume	Layer 1 = 96.6 cm
	Layer 2 = 106.6 cm
	Layer 3 = 116.8 cm
Radius	3.7 cm $<$ r $<$ 16.2 cm
Radial wire interval	4.57 mm
Number of cells/cell	8
Number of sense wires	640
Sense wire	25 μm NiCoTin, 1.8 kΩ/m 80 g tension
Sense wire potential	+2.5 kV
Guard wire	152 μm Au-plated Al
Gas mixture	CO_2(95%)-ethane(5%) at 1 atm
Gas gain	4×10^4
Drift Field	1.0-1.6 kV/cm
Drift velocity	7.3 μm/ns.

fields of the charged particle induce a time-dependent polarization in the medium, and this polarization emits radiation). At high energy, transition radiation is primarily emitted in the form of X-rays. These X-rays have an energy distribution that peaks at about 8 keV, and is mainly below 30 keV.

The TRD[41] is located between VTX and CDC, providing electron identification independent of that given by the calorimeters. The TRD consists of three separate units, each containing a radiator and an X-ray detection chamber. Radiators consist of 393 foils of 18 μm thick polypropylene, in a volume of dry nitrogen gas. The X-rays are detected in a radial-drift proportional wire chamber (PWC) mounted after the radiator, as shown in Fig. 2.8. Because of the contributions from transition radiation and the relativistic rise of the specific ionization[33], the amount of energy deposited in the TRD by electrons should be about twice the amount deposited by pions. Table 2.3 lists some parameters[40] of the TRD.

PWC

RADIATOR

exhaust

e⁻

π⁻

Xe+CH₄

CO₂

One module of the TRD
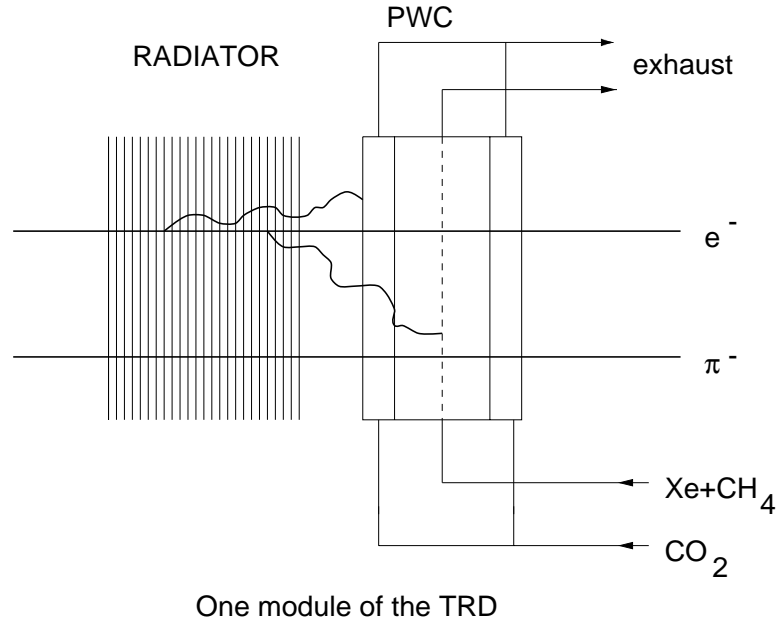
Figure 2.8: Schematic view of a TRD.

Table 2.3: TRD parameters.

Length of active volume	166.6 cm
Radius	17.5 cm $<$ r $<$ 49.0 cm
X-ray detector gas	Xe(91%), CH$_4$(7%),C$_2$H$_6$(2%)
Radiation length	0.0813 X$_o$
Interaction length	0.0357 λ$_o$
Gas mixture	xenon-CH$_4$
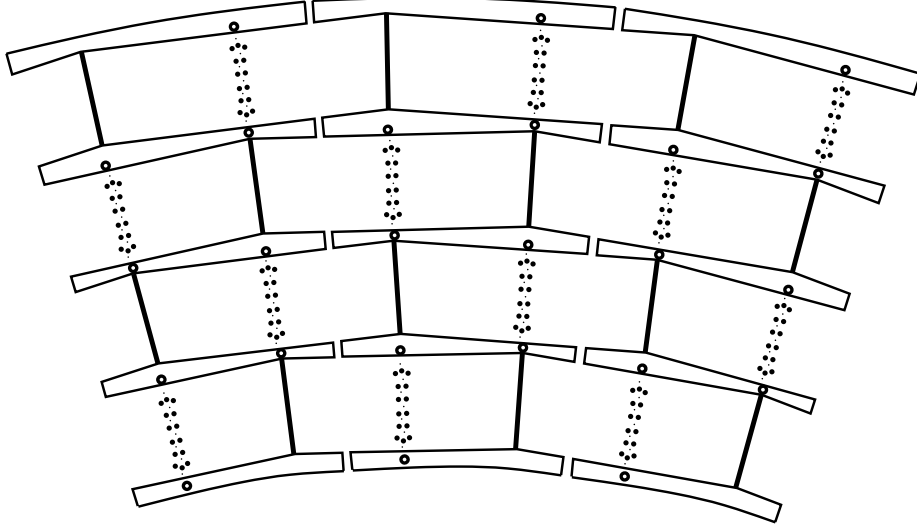
Figure 2.9: Schematic view of a CDC.

2.3.3 Central Drift Chambers

The central drift chambers provide tracking of charged particles, and their energy loss (dE/dx) for the pseudo-rapidity region $|\eta| < 1$. The CDC is a cylindrical shell of chambers 184 cm in length, with layers between of 49.5 and 74.5 cm (outside of the TRD). Figure 2.9 shows an end view of the CDC. Four different concentric rings contain 32 azimuthal cells per ring, and each cell contains seven 30 μm gold-plated tungsten sense wires. The CDC has a resolution in x,y of 180 μm and in z of 3.5 mm[37, 38]. Wire sizes, composition, and other details are given in Table 2.4.

2.3.4 Forward Drift Chambers

Forward drift chambers extend coverage for tracking of charged particles down to $\theta \approx 5^o$. The FDC occupies each end of the concentric barrels of the VTX, TRD, and CDC. As shown in Fig. 2.10, each FDC consists of separate chambers. Θ and Φ chambers measure θ and ϕ coordinates, respectively. Each Φ chamber is a single

Table 2.4: CDC parameters.

Length of active volume	179.4 cm
Radius	51.8 cm < r < 71.9 cm
Number of Layers	4
Radial wire interval	6.0 mm
Number of sense wires/cell	7
Number of sense wires	896
Number of delay lines	256
Gas mixture	$Ar(93\%)$-$CH_4(4\%)$-$CO_2(3\%)$-H_2O at 1 atm
Drift Field	620 V/cm
Drift velocity	34 mum/ns
Gas gain	$2,6 \times 10^4$
Sense wire	Au-plated W , diameter = 30 μm

chamber, containing 36 sectors over the full range of Φ. Each Θ chamber consists of four mechanically separate quadrants, each containing six rectangular cells at increasing radii.

2.3.5 Readout of the Central Detectors

The electronics for all CD devices consist of three stages of signal processing: Preamplifiers mounted directly on the chambers, the signal shaping electronics on the detector platform, and flash-ADC digitizers in the moving counting house (MCH). (The preamplifier output signals are carried out to the shaping circuits through 15 m long coaxial cables.) The full CD, including TRD detectors consist of 6080 separate channels.

2.4 Calorimeters

The energy of elementary particles can be measured with instruments that are generally called *calorimeters*[42, 43, 44]. A calorimeter is a block of matter in
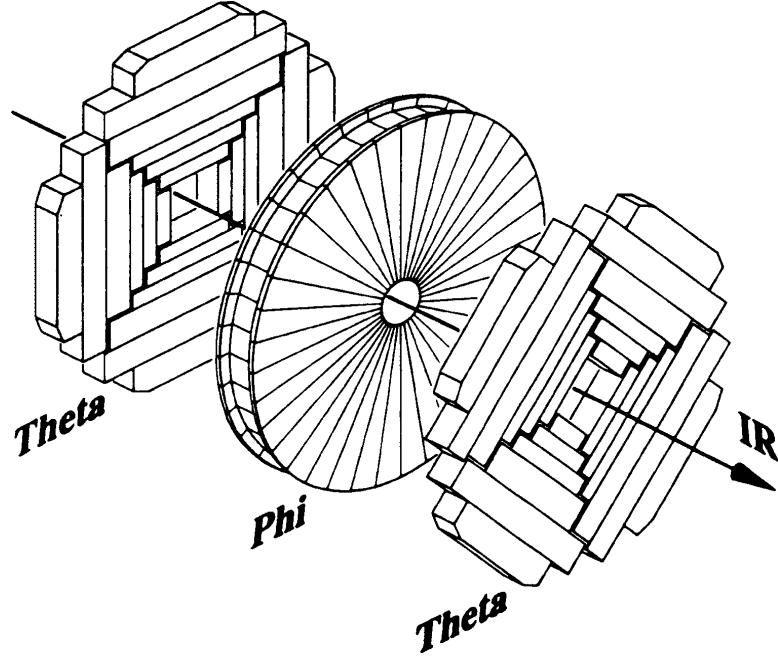
Figure 2.10: Schematic view of a FDC.

which the particle to be measured interacts and transforms part of its energy into a measurable quantity. DØ calorimetry is based on liquid argon for the sensitive region, and mainly uranium for the uninstrumented absorber.

The quantum of the electromagnetic field, the photon, interacts with matter via three different processes: the photoelectric effect, Compton scattering, and electron-positron pair production. Electromagnetic energy loss occurs through all three of these processes. A charged particle (e.g., an electron produced by an initial photon) can ionize the medium that it traverses, or it can radiate energy in the Coulomb field of a nucleus (Bremsstrahlung). If the incident photon's energy is large, then an electromagnetic shower develops in the medium. This is primarily characterized by the electron density in the absorber. To a certain extent, it is possible to describe shower characteristics in a material-independent way. For the longitudinal development, the so-called radiation length (X_o) characterizes the

shower size. The radiation length is defined as the distance over which a high energy electron loses on average 63.2 % (or $1 - 1/e$) of its initial energy. The Molière radius is defined by the ratio of X_o to ϵ_c, where ϵ_c is the energy at which the loss through radiation and ionization are the same (for an electron in the shower). This radius describes the transverse development of the shower.

When a high energy hadron penetrates a block of matter, it will at some point interact with one of its nuclei. In this process, mesons are usually produced (π, K, etc). Often, some fraction of the initial particle energy is lost in breaking the nucleus apart. If the nucleus is large, it can fission after absorbing very little energy, and emit nucleons and low energy photons. The higher-energy particles produced in the initial collision (mesons, nucleons, photons) can, in turn, interact again, or lose their kinetic energy by ionization. These processes also produce a shower. Such a hadronic shower is characterized by the nature of nuclear interactions, and the shower dimensions are governed by the nuclear interaction length, λ_{int}, which scales only approximately as the nuclear radius and more like $\sim 50 \times A^{1/4}$ (g/cm^2).

In any given calorimeter, the energy deposited by monoenergetic pions has a wider distribution than for electrons of the same energy. This is due to the fact that hadron showers suffer from larger fluctuations in their interactions. This is true for both the fraction of the total energy carried by ionizing particles as well as for losses to nuclear binding, which can consume up to 40% of the incident energy. The ratio of the electromagnetic (e) to hadronic (h) response of the calorimeter, e/h, should be close to unity, because otherwise the energy resolution for a complex shower (e.g., parton induced jet) is degraded.

The energy resolution of sampling calorimeters is usually dominated by the fact that the shower is sampled only periodically. The nature of such sampling fluctuations is purely statistical and, therefore, they contribute as $\sqrt{(E)}$ to the final energy resolution. An energy-independent noise term (electronics and radiation

from uranium), and a term proportional to the energy (e.g., from gain variations in amplifiers), also contribute to the energy resolution of the calorimeter system.

The DØ calorimeter provides the energy measurement of electrons, photons and jets, without the presence of a central magnetic field. In addition, it provides much of the information needed for identifying electrons, photons, jets, and muons, and plays an essential role in the determination of the transverse momentum balance in an event. Liquid argon was chosen as the active medium to sample the ionization produced in electromagnetic and hadronic showers. This choice was motivated by the desire of having a gain of unity (ionization chamber), by the relative simplicity of calibration and monitoring, by the good radiation hardness, and the low unit cost for readout electronics. However, liquid argon does have a complication, in that it must be operated in a cryogenic environment. Three cryogenic vessels were made to provide some degree of access to the central detectors within the calorimeter cavity (Fig. 2.11). The Central calorimeter (CC) covers roughly $|\eta| <$ 1, and a pair of end calorimeters (ECN (north) and ECS (south)) extend the coverage out to $|\eta| \approx 4$. An electromagnetic section (EM) with relatively thin uranium absorber plates, a fine-hadronic (FH) section with thicker uranium plates and a coarse-hadronic (CH) section with thick copper or stainless steel plates are the three distinct types of modules in both the CC and ECs.

A typical calorimeter cell is shown in Fig. 2.12. This generic unit cell consists of alternating layers of absorber and readout boards immersed in liquid argon. The readout board is a copper sheet (pads) sandwiched between two thin pieces of G10, which are covered with a resistive epoxy coating. The shower particles cross the liquid-argon gaps and ionize the argon atoms. The ionization electrons drift toward the resistive anode and induce a pulse on the copper readout pads.

The transverse sizes of the cells were chosen to be comparable to the transverse sizes of showers: $\approx$ 1-2 cm for EM showers and $\approx$ 10 cm for hadronic showers. The

DØ LIQUID ARGON CALORIMETER
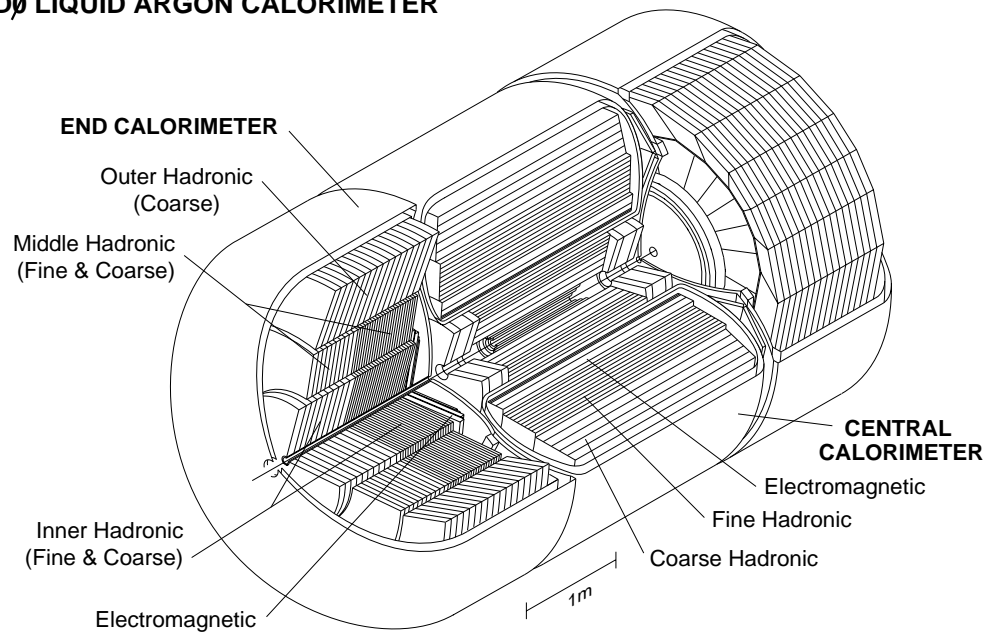
Figure 2.11: Isometric view showing the central and two end calorimeters.
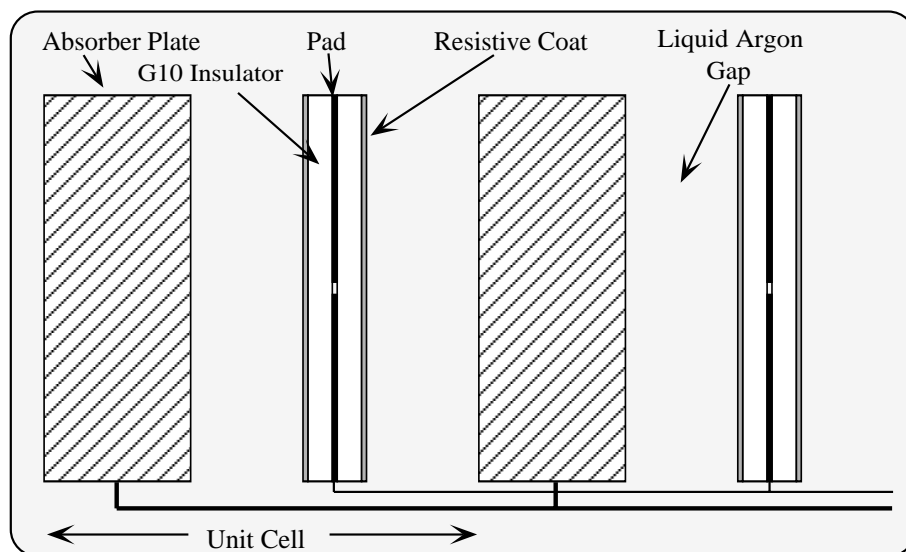
Figure 2.12: Schematic view of the absorber, liquid argon gaps, and signal board in a single calorimetric unit cell.
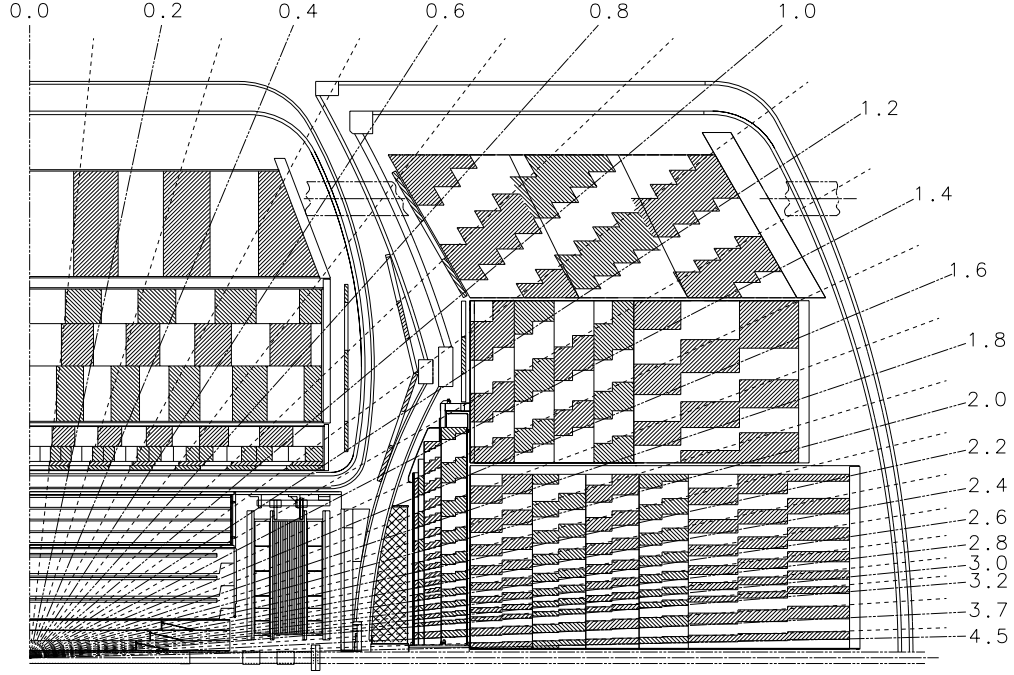
Figure 2.13: Schematic view of a portion of the DØ calorimeter showing the transverse and longitudinal segmentation pattern. Also shown are "massless" gaps at the end of the CC and the ICD in the region between the CC and EC cryostats.

readout boards are ganged together in a "pseudo"-projective geometry. Fig. 2.13 shows a portion of the segmentation pattern for the DØ calorimeter. There are four separate longitudinal layers for the EM in the CC and the ECs. The first two layers are about 2 X_o thick, and are used to gauge the longitudinal development of the shower, and thereby attempt to distinguish photons from π^os (in a statistical manner). The FH modules are segmented into three or four layers, and the CH modules are single or ganged into three layers. Typical transverse sizes of towers in both EM and hadronic modules correspond to $\Delta\eta = 0.1$ and $\Delta\phi = 2\pi/64 \approx$

Table 2.5: Central-calorimeter parameters.

	EM	FH	CH
Rapidity coverage	± 1.2	± 1.0	± 0.6
Number of modules	32	16	16
Absorber	Uranium	Uranium	Copper
Absorber thickness (inches)	0.118	0.236	1.625
Argon gap (inches)	0.09	0.09	0.09
Number of cells/module	21	50	9
Number of readout layers	4	3	1
Cells per readout layer	2,2,7,10	20,16,14	9
Total Radiation lengths	20.5	96.0	32.9
Radiation length/cell	0.975	1.92	3.29
Sampling fraction (%)	11.79	6.79	1.45
Segmentation[a] ($\Delta\phi \times \Delta\eta$)	0.1×0.1	0.1×0.1	0.1×0.1
Total number of readout cells	10386	3000	1224

[a] Third layer of EM has 0.05×0.05 segmentation

0.1. The third section in EM modules are twice as finely segmented in both η and ϕ to allow more precise location of centroids of EM shower[45].

2.4.1 Central Calorimeter

The central calorimeter covers the pseudo rapidity range of $|\eta| < 1.2$, as is shown in Fig. 2.13. It contains three concentric cylindrical shells of EM, FH, and CH modules. The CCEM modules have four longitudinal layers of approximately 2.0, 2.0, 6.8 and 9.8 X_o. A total of 32 modules along the ϕ direction in CCEM provide about 10400 channels, spanning 24 towers of $\Delta\eta = 0.1$ along the 260 cm length. The CCEM modules provide a precise energy and position measurement for electrons and photons in the central region. The CCFH modules have three longitudinal layers. and the CCCH modules contain just one segment. Table 2.5 shows some of the design parameters for the central calorimeters[37].

2.4.2   End Calorimeters

The end calorimeters cover the approximate region $1.1 < |\eta| < 4.5$. Each contains four types of modules. There is only a single EM and a single inner hadronic (IH) module, and no azimuthal cracks. Outside the EM and IH, there are concentric rings of 16 middle and outer hadronic (MH and OH) modules. The azimuthal boundaries of the MH and OH modules are offset to minimize particle penetration through cracks. The ECEM modules contain four readout sections of 0.3, 2.6, 7.9 and 9.3 X_o, with outer radii varying between 84 and 104 cm, and an inner radius of 5.7 cm. The two ECIH modules have inner and outer radii of 3.92 and 86.4 cm and have a cylindrical structure.

2.4.3   Intercryostat and Massless Gaps Detectors

A necessary but unfortunate design feature of the DØ detector is a large amount of uninstrumented material, in the form of cryostat walls, in the region of $0.8 < |\eta| < 1.4$, as indicated in Fig. 2.13. To correct for energy deposited in the uninstrumented walls, two scintillation counter arrays called intercryostat detectors (ICD) were mounted on the front surface of the ECs. Each ICD consist of 384 scintillator tiles of size $\Delta\eta = \Delta\phi = 0.1$. These are aligned with respect to the calorimeter, as indicated in Fig. 2.13. In addition, separate single-cell structures, called massless gaps, were installed inside both CC and EC calorimeters. One ring, of standard segmentation, was mounted on each of the end plates of the CCFH modules, and additional rings were mounted on the front plates of both the ECMH and the ECOH modules. The ICD and massless gaps provide sampling information in addition to that given by the standard DØ calorimetric sampling of showers.

2.4.4 Readout and Performance of the Calorimeter

Calorimetric signals are channelled to multilayer printed-circuit boards, and re-ordered from a module-oriented to η-ϕ oriented form, appropriate for subsequent analysis. The outputs are then brought to the preamplifiers, and are transported on 30 m twisted-pair cables to baseline subtracter shaping and sampling circuits. Input signals are integrated and differentiated. The main signals are sampled just before a beam-crossing and 2.2 μs after, and the difference is attributed to the collected charge. Subsequently, 24-channel 12-bit ADC circuits in the MCH digitize the sampled signals.

The observed energy resolution for electrons and pions has been parametrized as

$$\left(\frac{\sigma_E}{E}\right)^2 = C^2 + \frac{S^2}{E} + \frac{N^2}{E^2} \tag{2.1}$$

where C, S, and N represent the calibration errors, sampling fluctuations and noise contributions respectively. For electrons, the measured resolutions are[32]:

$$C = 0.003 \pm 0.002, S = 0.157 \pm 0.005\sqrt{\text{GeV}}, N \approx 0.140 \text{ GeV} \tag{2.2}$$

and for pions,

$$C = 0.032 \pm 0.004, S = 0.41 \pm 0.04\sqrt{\text{GeV}}, N \approx 1.28 \text{ GeV}. \tag{2.3}$$

2.5 Muon Detectors

Muons usually provide one of the cleanest signals in the collider environment. Because muons are approximately 200 times heavier than electrons, for energies below about 500 GeV, they only rarely produce electromagnetic showers

Table 2.6: Parameters of the muon system.

Toroids	Central	End(each)	SAMUS(each)
Z interval (cm)	±378.5	447.0-599.4	447.0-599.4
Inner distance (cm)	±317.5	±91.4	±25.4
Outer distance (cm)	±426.1	±416.6	±85.1
WAMUS			
Total no. of chambers	164		
Maximum length	579 cm		
Maximum width	254 cm		
Total no. of cells	11386		
Total no. of wires	11386		
SAMUS			
No. of planes/station	3		
Total no. of planes	$2 \times 3 \times XYU = 18$		
Size of planes	330 cm$\times$330 cm		
No. of wires per plane	$256(X, Y)$, $360(U)$		
Total no. of channels	5376		
Resolution	$\sigma = 0.2$ mm		
Gas mixture	$ArCO_2$ or $ArCF_4$		
Drift time	200 ns (Ar + 10% CO_2)		
	160 ns (Ar + 10% CF_4)		

(Bremsstrahlung will occasionally initiate an EM shower). Muons also do not have strong interactions. Consequently, they leave minimum-ionizing tracks, and can be identified even in the middle of hadron jets.

The DØ muon detection system[46] consists of solid-iron toroidal magnets, together with sets of proportional drift tube chambers (PDTs) that measure the track coordinate and its momentum down to 3°. Figure 2.14 shows an elevation view of the DØ detector with the five toroids and their associated PDT layers indicated. The central toroid (historically, the central toroid has been called CF, referring to central fero-) covers the region $|\eta| < 1.0$, and the two end toroids (EF) cover $1 < |\eta| < 2.5$. The Small angle muon system (SAMUS) toroids fit in the central hole of the EF toroids, and cover $2.5 < |\eta| < 3.6$.

SAMUS PDT

A STATION
B STATION
C STATION

WIDE ANGLE PDT (CM)
A LAYER
B LAYER
C LAYER

WIDE ANGLE PDT (EM)
A LAYER   B LAYER   C LAYER

90°   80°   70°   60°   50°   45°   40°   35°

30°
25°
20°
15°
10°
5°

CF
EF
EF

CENTRAL
DETECTOR

QUAD

QUAD

0°

CALORIMETER

EC   CC   EC

TOROID   CF

PLATFORM &
ELEC. RACKS

SAMUS TOROID

(feet)
0   5   10

(meter)
0   1   2   3
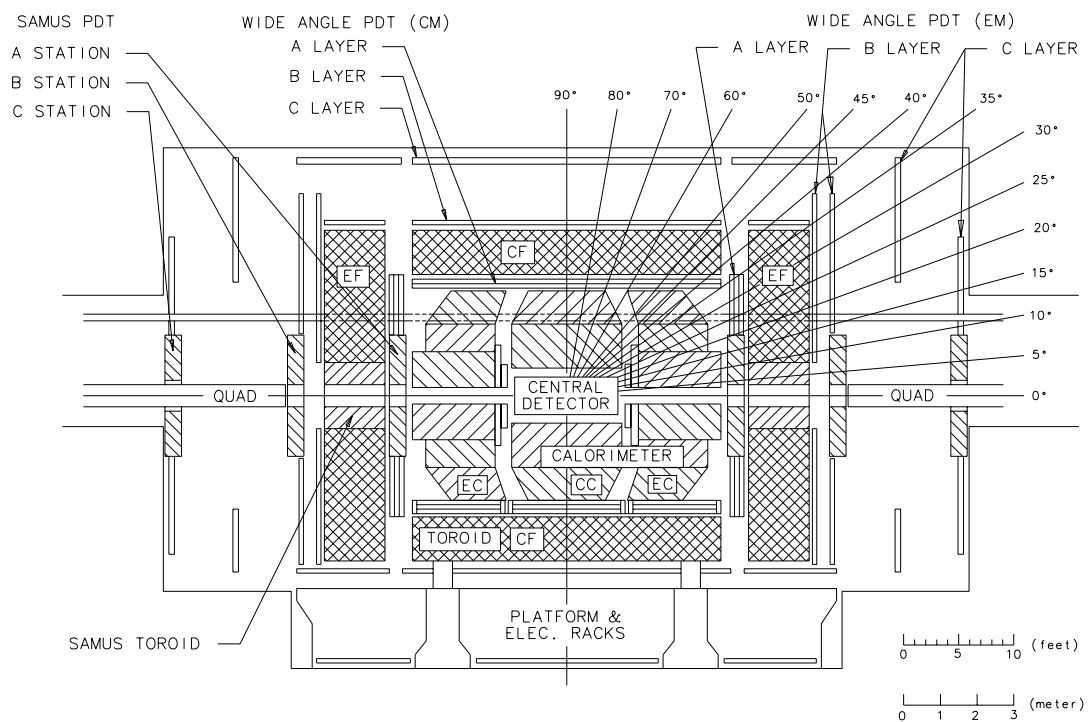
Figure 2.14: Elevation view of the DØ detector showing the five toroids and the approximate dispositions of the A,B, and C layers of proportional drift tubes.

2.5.1 Muon Toroids

The CF toroids have twenty coils of 10 turns each, and carry currents of 2500 A. This provides an inner magnetic field of 1.9 T. The two EF toroids are located at $447 < |z| < 600$ cm. The EFs comprise a 183 cm square. Eight coils of eight turns each also carry 2500 A, thereby providing fields of about 2 T.

2.5.2 Muon Chambers

The wide-angle muon system (WAMUS) PDTs are constructed from aluminum cells, as shown in Fig. 2.15. The "A" layer, before the iron toroids, has four cells, and the "B" and "C" layers after the magnets, each have only three layers of cells. Cathode-pad strips are located at the top and bottom of each cell, and an anode wire (marked $\times$) is held near the center of the cell. The coordinate (ξ) along the wire direction is measured through a combination of cathode pad signals and timing information from the anode wires. The ξ resolution is approximately $\pm$ 3 mm.

The A layer of the SAMUS system is located before the SAMUS toroid; the B and C layers are between the toroid and the beginning of the low-beta quadrapole for the DØ insertion. The SAMUS PDTs consist of 3 cm external diameter stainless steel tubes with individual end plugs for gas and electrical connections. There are a total of 5308 tubes in the SAMUS system. A list of muon-system specifications can be found in Table 2.6.

2.5.3 Readout and Performance of Muon Chambers

Much of the signal-processing electronics for the PDTs resides on the chamber modules, digitizers, and trigger electronics reside in the MCH. Signals from each cell of WAMUS are brought to a charge-sensitive preamplifier that is similar to the
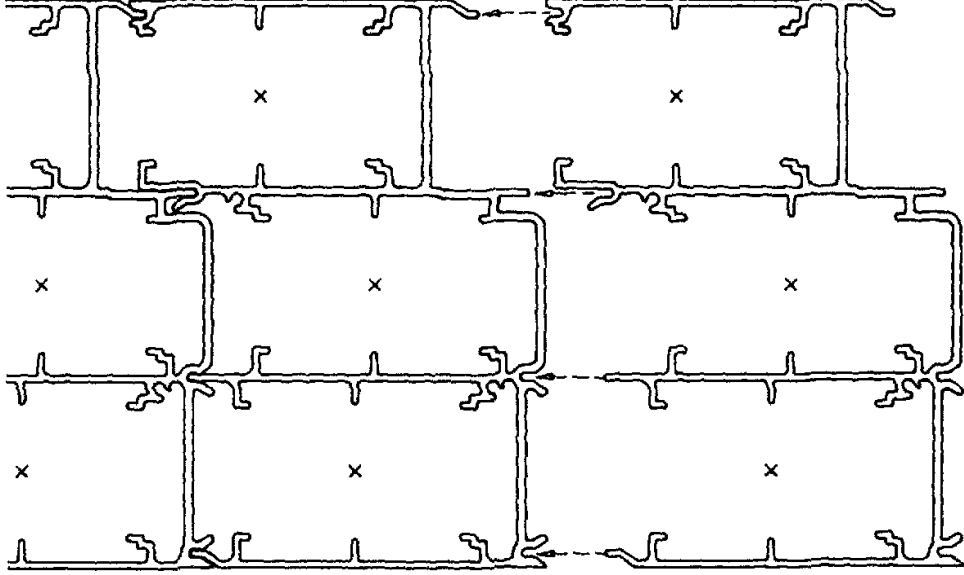
Figure 2.15: Extruded aluminum section from which the B and C layers of PDT chambers are constructed. The A layer extrusions are similar, but have four cells instead of three. The × signifies positions of anode wires. Cathode strips are inserted at the top and bottom of each cell.

one used in the calorimeter. Digitization of the signals from the chamber cathode pads and determination of timing information is performed in the MCH using a 12-bit ADC circuit.

PDT drift-coordinate resolution is about ± 0.53 mm. Studies of the chamber efficiency as a function of position within the unit cell showed nearly full efficiency. Cosmic ray studies were also showed that the resolution for the ξ coordinate is ± 3 mm.

2.6   Triggering and Data Acquisition

The DØ trigger and data acquisition systems[47] selects the few interesting events to be recorded from typically 5×10^5 $\bar{p}p$ interactions/sec. The trigger systems have four distinct levels: Level 0, which is formed by coincidences of

elements of scintillator hodoscopes on either side of the interaction region; Level 1, comprised of muon and calorimeter triggers; Level 1.5, having only the TRD and several components of the muon trigger as inputs, and, Level 2, which is based on a large number of parallel microprocessors that analyze individual events. For Levels 0 and 1, the trigger is deadtime-less, as decisions are made in the time between successive bunch crossings so that no events go unexamined. This is no longer true for Level 1.5, which needs tens of microseconds for its analysis, thereby exceeding the time between beam crossings, even with only six bunches of protons and antiprotons in the Tevatron.

2.6.1   Level 0 Trigger

The Level 0 system[48] is designed to register the presence of an inelastic collision, to provide a fast estimate of the location of the event vertex, and to serve as the luminosity monitor for the experiment. It consists of two separate hodoscopes of scintillation counters located at each end of the central detector between the FDC and the EC. The signals are read out with photomultipliers. The rapidity coverage is partial for the range $1.9 < |\eta| < 4.3$ and nearly full for the range $2.3 < |\eta| < 3.9$. A coincidence of both Level 0 detectors is estimated to be about 99% efficient in detecting non-diffractive inelastic collisions[48]. The z coordinate of the primary collision vertex is provided by comparing the arrival times of the signals from the two scintillator arrays. Time resolution of each of the Level 0 counters is in the range of 100 - 150 ps which provides the vertex position to about $\pm$ 2 cm[48]. At an instantaneous luminosity of 5×10^{30} events cm^{-2} s^{-1} there are on average of 0.75 interactions per crossing, and for the case of multiple interactions, the Level 0 system sets a flag that is used in the subsequent trigger levels. The Tevatron luminosity is monitored by measuring the rate for

non-diffractive inelastic collisions, and this information is sent back to accelerator operations for use as a feedback mechanism[48].

2.6.2 Calorimeter Trigger for Level 1

At its readout level, the DØ calorimeter consists of pseudo-projective towers with segmentation of 0.1 in both ϕ and η, with seven or eight subdivisions in depth. For the trigger, the readout towers are added laterally 2 by 2 to form trigger towers with transverse dimensions of $\Delta\phi = 0.2$, $\Delta\eta = 0.2$. In depth, all the EM sections are added to form an EM trigger tower, and the corresponding hadronic trigger tower contains the remaining section except the last one. The very last or coarse hadronic sector is excluded because it generally contributes more noise than signal to the various sums.

There are two broad categories of triggers that can be constructed using energies deposited in individual calorimeter towers: "Global" triggers that use quantities such as the transverse momentum (or energy) deposited in the EM sectors, or in the hadronic sectors, or in the full towers. The other major category of triggers, or "cluster triggers", are derived from considering the "transverse energy" deposited in individual EM and in the sum of EM and hadronic towers. The number of trigger towers with deposited transverse energy in excess of any of four preset thresholds is counted, and the resulting counts are then compared to as many as four limits for each energy threshold, and used to generate input terms for AND-OR network.

2.6.3 Level 1 Muon Triggers

The Level 1 muon trigger modules[49] consist of VME-based cards designed to find track segments in individual chambers, and then match each other. Each of the muon triggers generate bit patterns corresponding to hit centroids for the
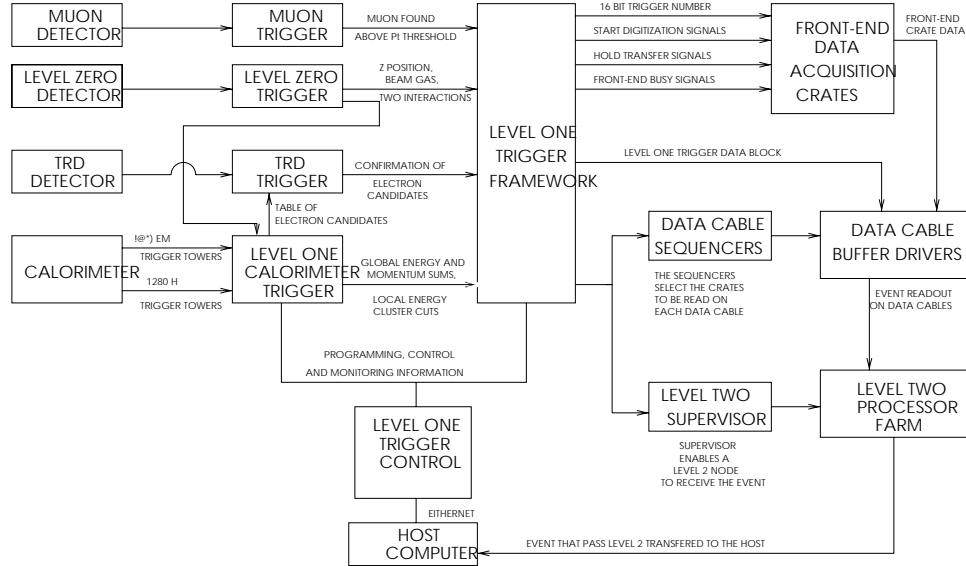
Figure 2.16: A diagram of the DØ trigger indicating individual trigger components and their interconnections.

Level 1 and 1.5 trigger electronics. Comparison of bit patterns in A,B and C layers determines a good Level 1 muon trigger.

2.6.4 Level 1 Framework

The heart of the DØ trigger system is the Framework. As shown in Fig. 2.16, the Framework accepts inputs from various devices that contribute to the trigger decision and, chooses events of interest for further processing. In addition, it coordinates the various vetos that can inhibit triggers, provides a prescaling function for each trigger, manages the communication tasks with the front-end elements and with the Trigger Control Computer (TCC), correlates trigger and readout functions, and is the repository of the large number of scalers that are essential for tracking such quantities as trigger rates and deadtimes.

- Main Ring Veto

  As is described in the previous section, the Main Ring passes through the

hadronic section of the DØ calorimeter and through the EF muon chambers. Because of this, any losses from the Main Ring can produce backgrounds in the detector and must be removed. The injection of the beam in the Main Ring, and the "transition" after 0.3 seconds after injection, cause major losses[32]. These losses are vetoed with the help of a trigger component MRBS_LOSS (MRBS referring to Main Ring Beam Synchronous Clock)[51]. This introduces a veto during a 0.4 seconds window of the injection time of beam into the Main Ring. This causes a dead time of about 0.4/2.4, or $\sim$ 17 %[51]. (The Main Ring cycle is 2.4 seconds.) There is also a possible loss when a bunch in the Main Ring passes through the detector. Due to this, an additional trigger component, the MICRO_BLANK, is implemented. This veto is set for any beam crossing when a Main-Ring bunch is present in the detector within $\pm$ 800 ns to the experiment of the crossing. Imposing this option, this term adds an additional $\sim$ 8 % deadtime[52].

Both of the above vetos were active for the first part of the data taking (RUN IA), and a more efficient scheme was developed[52] for the rest of the data taking (RUN IB) and increased the live time about 8 %[52].

2.7 Data Acquisition

About 1 ms after receipt of a valid Level 1 or 1.5 trigger, fully digitized data appear in the output buffers of the approximately 80 VME crates containing the calorimeter and muon chamber ADCs and the tracking and TRD chamber flash ADCs. Each crate of primary digitized data contains a 512 kBytes memory module with two data buffers. The buffer outputs for each particular sector of the detector are connected sequentially to a high speed data cable. The data cables

circulate to each of the 50 Level 2 processor nodes that lead directly into the desired unprocessed data with ZEBRA[50] structure.

2.7.1 Level 2 Filter

Event-filtering in the 50 Level 2 nodes reduce the approximately 100 Hz of input rate to 2 Hz, which can be logged for offline analysis. The filtering process in each node is build around a series of software tools. Each tool has a specific function related to identification of particles or event characteristics, such as the presence of jets, muons, EM clusters, tracks associated with a calorimeter cluster, $\sum E_T$, and missing E_T. The Level 2 nodes are coordinated through the host computer, and have access to current distributions of parameters and statistics on recent processing history characteristics.

2.7.2 Host Computers

The host cluster consists of a VAX 6620, VAX 6410 and VAX 8810 processors, and a set of shared disks, together with VAX stations connected by an Ethernet/FDDI network. The 6620 is the primary data collection engine, receiving events from the Level 2 output data-cable. The 6410 is the primary machine responsible for spooling events from the staging disk to 8 mm tape. The 8810 is devoted primarily to monitoring hardware. The host computer also produces the primary human interface to the detector systems, and is responsible for high-level control of the data-taking system, downloading of all settable parameters, specifying hardware monitoring activities, and the recording and displaying of data obtained with the detector.

Chapter 3

Reconstruction

3.1 Offline Data Processing

The unprocessed data taken by the online system of the DØ detector (commonly called "raw" data) and the detector constants are used as input to the offline DØ event reconstruction program (DØRECO), which identifies the vertex position, trajectories of charged particles, energy deposition by electromagnetically interacting objects and by hadronic jets, and transverse momentum balance in events. The reconstruction at DØ is performed on a farm of 74 Silicon Graphics Indigo (SGI 4D/420) nodes and 24 IBM 220 nodes operating in a UNIX environment. During reconstruction, groups of data describing an object or characteristics of given event, or *bank*s, are created, linked, and either passed on or dropped. Selected events are written out to two types of files: the standard output (STA) and the data summary tapes (DST). The STA files contain all the information that is necessary for event re-reconstruction, including HITS bank (signals collected from all detectors). The DST files are compressed version of the STA files containing only processed information, and, in particular, no information from the HITS bank.
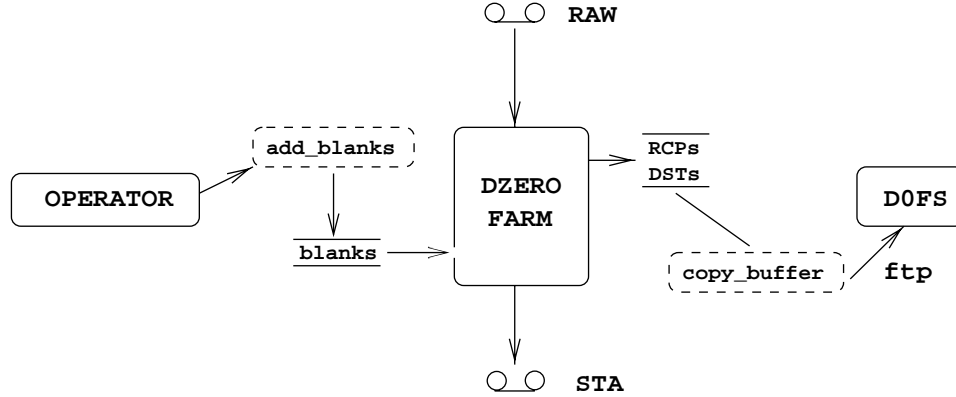
Figure 3.1: A block diagram showing the data flow in offline processing.

After DST files are finally transferred to the DØ file server (DØFS), more information is dropped and compressed to fit on the storage disk. The latter is called compressed-micro DST. Figure 3.1 shows the overall flow of the data stream for the entire offline data processing. The entire process, starting with the reading of unprocessed data from tape, to the writing of micro DST files to DØFS, is quite complex and requires substantial human intervention.

3.1.1 Farm Hardware

The DØ farm consists of UNIX workstations and servers, configured logically into two different sets. Most of the workstations (nearly 100 nodes) are used as "worker nodes" that perform the reconstruction of unprocessed data: each consists of a UNIX workstation with 24-32 MBytes of memory and a local system disk, with ethernet and power connections. These workstations are a mixture of 6 SGI 4D/35, 68 SGI R3000 Indigo, and 24 IBM 220s. The combined CPU power is comparable to approximately 3000 MIPS (million instructions per second). The other set of workstations are "servers", that is, designated as I/O servers that supply the unprocessed data to worker nodes, and spools input (output) files to local disks (8

mm tapes). DØ uses four SGI 4D/420 I/O servers.

The balance between I/O servers and worker nodes must be tuned to maximize the available computing power. The average ratio of worker nodes to an I/O node is approximately 16 - the ratio used for any given application is determined primarily from experience and the CPU and I/O needs of a problem[53]. To avoid saturating ethernet segments, the worker nodes are divided into subnets attached to routers, with each subnet consisting of between 8 and 10 workstations. Figure 3.2 shows the configurations of worker nodes and servers. The fnsfxxx are the hostnames for SGI Indigo worker nodes, and fnckxxx are the names for IBM 220 worker nodes (with xxx representing numbers in Fig. 3.2). Worker nodes are grouped into four "farmlets". Each farmlet consists of an I/O node (SGI 4D/420, denoted as fnsfb, fnsfd, fnsfe, and fnsff) with 3 ethernet interfaces and 24 worker nodes. (This is the definition of a farmlet used by Fermilab farm group [53]. However, DØ uses this word to describe any group of worker nodes communicating with a given I/O server, such as those grouped inside of the boxes in Fig 3.2). The size of the executable DØRECO program constrains each worker node to at least 24 MBytes of real memory (some, such as fnsf159 - fnsf166 and all IBM worker nodes, have 32 MBytes). Groups of 8 or 10 worker nodes (in Fig. 3.2) reconstruct DØ events with the given executable, and read (write) input (output) files from (to) a common spool disk. A parallel server process running on an I/O node is called a "virtual machine" (VM), which corresponds to a logical machine that controls the group of 8 or 10 worker nodes. The generic name fnsfX_Y is used to identify a specific virtual machine. Capital X is one of b, d, e, f and Y is one of 0, 1, 2. Although VMs do not exist physically, it is common to designate the worker nodes associated with an I/O server in terms of their VM. Each I/O server has 7 tape drives for reading (writing) unprocessed (reconstructed) data from (to) 8 mm tapes. Failures with tape drives (SCSI reset, tape stuck in the drive, and

Figure 3.2: A block diagram showing the configuration for the farm hardware.

bad tapes) are one of the major reasons for downtime on the DØ farm.

3.1.2 Farm Software

Farm software controls the data flow at each step of event reconstruction. It requests tape mounts to operators, spools the unprocessed data, distributes events to worker nodes, writes the reconstructed data to spooling disks, merges output files when necessary, writes output data to tapes, transfers DST files to DØFS, and monitors all individual processes. The software is written in a combination of c-shell scripts, PERL (Practical Extraction and Report Language) scripts, FOR-TRAN, and C language. A block diagram of the software components is given in Fig. 3.3.

Primary (unprocessed) data tapes written by the DØ online system are transferred to the Fermilab Computing Center (FCC), and are stored for eventual reconstruction. A process running on each farmlet, the inspooler[54], checks the space on spooling disks, clears up leftovers from the latest failure, identifies the next primary data tape, and finally submits a data tape mount request to the operators at FCC. The inspooler contains a manager package written mainly in PERL scripts. The part that requests tape mounts is written in FORTRAN and uses OCS[55] (operator communications software) for management of tape drives. A tape mount involves a use of RBIO[56] (Raw Buffered I/O) for reading VMS labeled tapes. Once a data tape is mounted, the inspooler spools data to the spool disk (for example, /spool00/dzero/fnsff_0/inspool/), each specified for different VMs. This inspooling process is performed on each I/O server, without the participation of worker nodes. Log files from the inspooler indicate that the spooling speed is on the order of ~ 300 Bytes/s, where the time spent waiting for tape mounts or waiting for free disk space is not included.
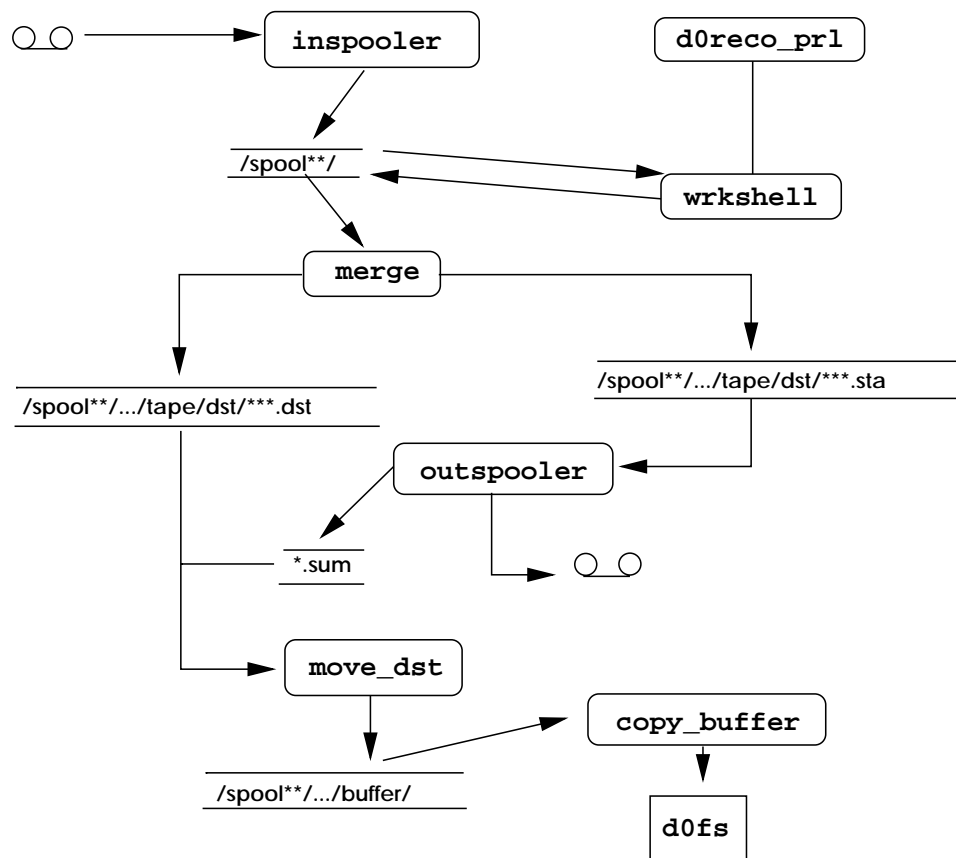
Figure 3.3: A block diagram showing the configuration for the farm software.
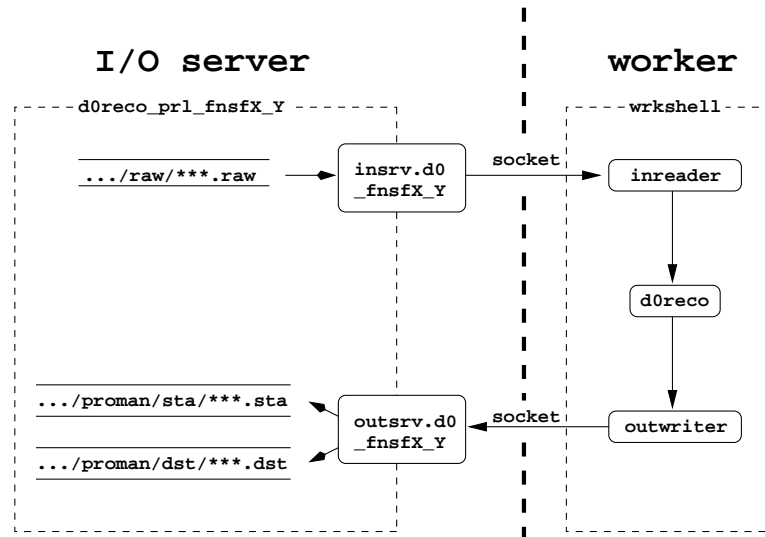
Figure 3.4: A block diagram showing details of the prl_reco process.

d0reco_prl initiates transmission of events to and from worker nodes. When d0reco_prl recognizes the presence of unprocessed data in the spool disk, it starts wrkshell that executes the DØRECO program on worker nodes. Output files (STA and DST files) are spooled to the disk on I/O servers. Figure 3.4 shows the data flow between d0reco_prl and wrkshell. Typically, the sizes of files are $\sim$ 180 MBytes/file for STA and $\sim$ 30 MBytes/file for DST. On each worker node, a process called inreader receives events from the server node and stores them in a shared memory location, where they can be accessed by the DØRECO program and by the outwriter for transfer of events back to the outsrv process after reconstruction.

Next, merge produces RCP (Run Control Parameter) files for DST and STA files. Only for the case of the DØ FIX project[58] are DST files truly merged, (10 files to 1 file). Once the merge process produces RCPs, outspooler starts spooling STA files to 8 mm tapes, using techniques similar to those used in the inspooler process. If outspooler finishes the outspooling of a tape, then the mvdst process

moves DST files and corresponding RCP to a buffer area. On the average, ~ 20 - 40 STA files are written per tape. Finally, RCP_move moves STA-RCPs to DØFS, and copy_buffer transfers DST files with their corresponding RCP files to DØFS.

The continuous operation of the farm requires access to types of 8500 and 8505 tape drives. Because the aging of tape drives leads to failures, it is important to keep track of files that are not successfully completed on a first attempt. For this reason, a database is maintained on the progress of files through various stages of the system, which provides the possibility for resubmission of missed partitions on a weekly basis.

Monitoring of processes on the farm is an independent operation from the existing parallel processes. mon_disp displays the current status of the utilization of the farm, and helps identify the presence of any problems.

3.1.3 DØ File Server

DØFS consists of four DEC 3000 AXPs and several DEC VAXstations, operating under VMS. The AXPs serve as disk file servers and have 300 GBytes of SCSI based disks attached to them. The VAXstations are employed as tape-server nodes, and are equipped with 130 GBytes disks and 31 of 8 mm tape drives. Based on the RCP files transported from the farm, STA and DST files are catalogued and distributed using the File and Tape Management (FATMEN) system[57] developed at CERN. The FATMEN package provides access to file catalogue information and the file itself, without requiring prior knowledge of the location of the file.

3.2 Reconstruction of Objects

The reconstruction of particle trajectories and other such objects is performed by the DØRECO program. The DØRECO program consists of hooks (an interface or logical function that performs an individual task) that reconstruct a primary vertex, charged tracks, jets, photons, electrons, muons and the transverse momentum imbalance, etc. We will now review some of these tasks.

3.2.1 Vertex Determination

Before executing DØRECO, the (x,y) position of the interaction point is determined by an online program called CD-EXAMINE[59]. Consequently, DØRECO is used to determine only the z-coordinate of the interaction point(s). This is done using mainly tracks in the CDC. All CDC tracks are extrapolated to the z axis, and the intersection in z for each track is stored. A cluster-finding algorithm is used to determine the number of clusters associated with any given z position. Once this is done, a constrained fit is performed to yield a precise measurement of the z vertex. For an event with multiple interactions, up to three possible candidate vertexes can be defined.

3.2.2 Identification of Jets

The first thing that is done to reconstruct jets (or for any other tasks requiring calorimeter information) is to convert the ADC values of each cell in the calorimeter to an equivalent energy. The conversion formula is:

$$E(e,p,l) = A(d) \times W(e,l) \times C(e,p,l) \times G(e,p,l) \times ADC(e,p,l) \qquad (3.1)$$

where e is the detector η-index (-37 $< e <$ 37), p is the detector ϕ-index ($1 \leq p \leq$ 64), l is the detector depth index ($1 \leq l \leq$ 17), d is the type of module (CC for central, EC for end cap, ICD, CCMG, or ECMG), A is an overall calibration constant (different for each d), W are sampling-fraction weights (determined from test-beam data), C is a run-independent correction (for absorber thickness or for signal in the ICD), G are run-dependent gain corrections, ADC is the ADC count observed in any given cell of the calorimeter. Once this conversion is done, then jet algorithms can be applied to the signal.

Roughly speaking, a jet is a localized cluster of energy deposited in the calorimeter, and that originated from scattered partons in collisions of protons and antiprotons. However, because of color confinement and gluon radiation from partons, the exact definition of a jet often depends on the physics process one wants to study. For this reason, DØ uses more than one jet definition. A fixed cone algorithm[60] has been the primary choice for defining jets in DØ. This choice has the advantage of straightforward applicability in theoretical calculations, in Monte Carlo simulations, and to data. A secondary choice for defining jets is a nearest-neighbor algorithm[61], in which neighboring cells are joined to form jet clusters, depending on their relative proximity and energy. This procedure is based on a local equivalence principle[61].

- Cone Algorithm

  The energy vector $\mathbf{E_i}$ associated with a calorimeter cell i is defined as the scalar energy E_i directed from the interaction point to the center of cell i. The tower vector $\mathbf{E_k^{tower}}$ is defined as the vector sum of the $\mathbf{E_i}$ over a semi-projective tower k. The transverse energy of the tower, E_T^{tower} is then given as:

  $$E_T^{tower} = E^{tower} \frac{\sqrt{(E_x^{tower})^2 + (E_y^{tower})^2}}{\sqrt{(E_x^{tower})^2 + (E_y^{tower})^2 + (E_z^{tower})^2}}. \qquad (3.2)$$

Jet finding in DØ is based on these tower E_T values. The algorithm is implemented in a three-step process. The first step involves preclustering of towers with E_T greater than some threshold value. Second, a jet energy and a jet axis are calculated from the sum of preclusters within a cone of some radius $\mathcal{R}$ in (η, ϕ) space. The jet axis is recalculated using these towers, and the process is iterated until the jet axis moves a distance less than 0.01 in (η, ϕ) space between two final iterations. Finally, based on the closeness of two jets in (η, ϕ), splitting and merging of jets can be performed.

The final kinematic quantities defining a jet are:

$$E_i = \sum_{towers\ k} E_i^k, \tag{3.3}$$

$$E_T = \sum_{towers\ k} E_T^k, \tag{3.4}$$

$$\phi = \arctan(E_y/E_x), \tag{3.5}$$

$$\theta = \arccos(E_z/\sqrt{E_x^2 + E_y^2 + E_z^2}), \tag{3.6}$$

$$\eta = -\ln\tan(\theta/2), \tag{3.7}$$

where i represents the four components of the four-momentum vector. Note that E_T (just as E) is the scalar sum of the transverse energies of the individual towers, and not the magnitude of their vector sum. Thus large E_T corresponds to the release of a large amount of energy in the collision, independent of the questioning of balance of transverse momentum.

- Nearest-Neighbor Algorithm
  The clustering logic used for this algorithm is identical to that used for electrons and photons, except that the parameters are optimized for jet finding. As before, the clustering starts with the calorimeter towers. For each tower, the neighborhood in η-ϕ (defined by the 24 towers surrounding the origi-

nal tower) is searched for the tower with the highest E_T. This E_T tower is *connected* to the original tower if its E_T exceeds a given threshold. The connection is performed until no towers remain unconnected. Such connections are final, and neither splitting or merging is subsequently allowed.

The performance of the jet-finding algorithms is measured primarily by the fraction of partonic jets that are reconstructed successfully (or jet reconstruction efficiency, referred as ϵ). Figure 3.5 shows reconstruction efficiencies of various jet-finding algorithms for HERWIG $t\bar{t}$ events (only multi-jet final states). The definition of a quark in the figure includes any final-state radiation added back to the quark momentum. The matching of reconstructed jets to quarks relies on using combinations that minimize the distance in $\mathcal{R}$ between them. A jet is considered to be matched only if $\Delta\mathcal{R} < 0.5$, and the energy of the jet is within a factor of two of the quark energy. No jet with reconstructed $E_T < 10$ GeV is considered. Figures 3.5 (a) and (b) show how the reconstruction efficiency depends on quark E_T and on quark η for the nearest neighbor algorithm ($\bullet$), and for the $\mathcal{R}$=0.3 (dashed line), $\mathcal{R}$=0.5 (dotted), and $\mathcal{R}$=0.7 (dot-dash) cone algorithms. The nearest-neighbor and 0.3 cone algorithm show similar level of jet reconstruction efficiency, but the 0.5 and 0.7 cone algorithms are less efficient. (The nearest neighbor and 0.3 cone algorithms provide an efficiency of $\approx 95\%$, but the 0.5 cone is about 92%, and the 0.7 cone is about 84% efficient at $|\eta| \sim 0$.) Cone algorithms for large cones are less efficient in a multi-jet environment ($t\bar{t}$ production usually provides five or more jets in $|\eta| < 2$ for multi-jet final states). Figure 3.5 (c) shows the correspondence between parton and jet energies found for various algorithms, using the DØ jet-energy correction package, CAFIX version 5.0 (see later in this section). For the case of cone algorithms, only linear fits to the correlation are shown in the figure.

The overall efficiency, or event reconstruction efficiency, for finding $t\bar{t}$ events

with 6 jets scales, of course, as ϵ^6. Figure 3.5 (d) shows the three-jet invariant mass for the correct combinations of jets matching top and antitop quarks. The area for the mass distribution reflects the $t\bar{t}$ event reconstruction efficiency for different algorithms. The nearest neighbor algorithm gives the highest efficiency, and the efficiency decreases as the cone size increases (44% for the nearest neighbor algorithm, 38% for $\mathcal{R}$=0.3, 31% for $\mathcal{R}$=0.5, and 18% for $\mathcal{R}$=0.7 cone algorithms). The shift of reconstructed mass from the input mass of the top quark (160 GeV/c^2) shows that the energy scales used for each of the jet algorithms are not equivalent. The $\mathcal{R}$=0.3 cone scale corresponds to a shift 13% in three-jet mass, $\mathcal{R}$=0.5 corresponds to 9%, $\mathcal{R}$=0.7 corresponds to 7%, and the nearest neighbor corresponds to 2%. The RMS values of the mass distributions are not affected very much for the different algorithms, providing $\sim$ 10% as an overall spread in reconstructed mass. The ability to resolve separate jets is important, not only in the search for a signal, but also for mass measurement. The jet algorithm is the fundamental tool in these analyses, and either the nearest-neighbor algorithm or the $\mathcal{R}$=0.3 cone algorithm appear to be preferred for resolving multi-jet events. However, due to the relatively large uncertainties in the jet energy scale of both the 0.3 cone algorithm and the nearest-neighbor algorithm, we chose to use the 0.5 cone algorithm as the primary algorithm to analyze multi-jet events in this dissertation.

3.2.3 Corrections to Jet Energies

For a variety of reasons, such as inefficiency of reconstruction algorithms, detector imperfections, overlap with energy from fragmentation of spectator partons, etc., the energies of reconstructed jets can be biased. To correct for this, DØ developed a procedure, referred to as the CAFIX package[62, 63]. The underlying assumption in the correction is that the true energy of a jet (E_{true}) can be ob-
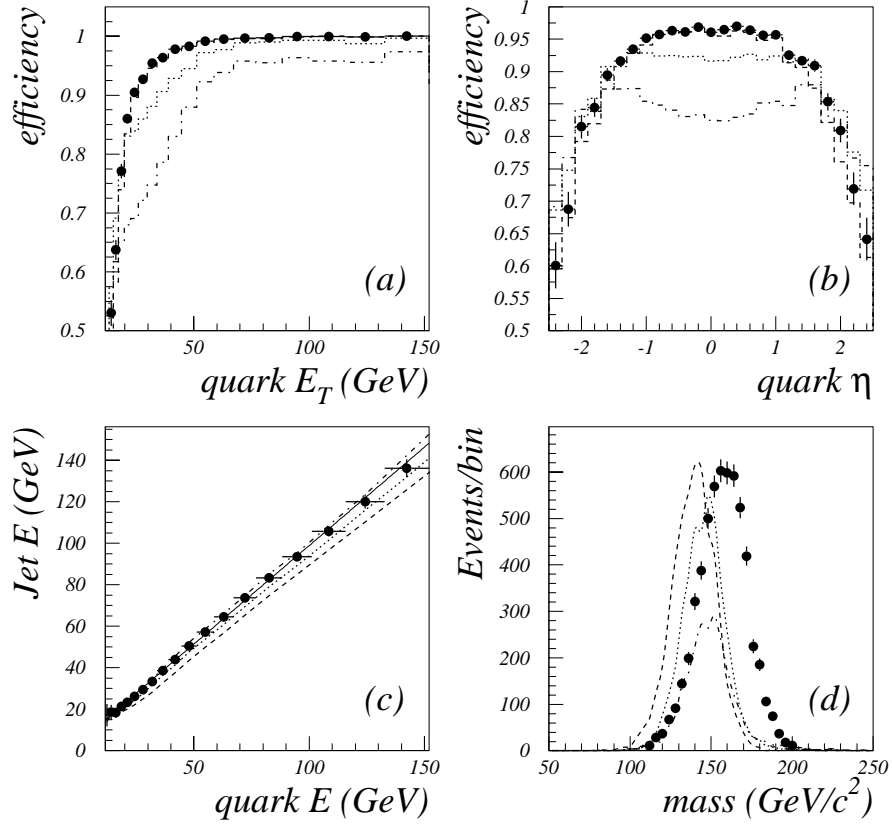
Figure 3.5: Jet finding efficiency for $t\bar{t}$ events ($m_t = 160$ GeV/c^2) with various algorithms: nearest neighbor=($\bullet$), cone algorithms of $\mathcal{R}$=0.3 (dashed line), $\mathcal{R}$=0.5 (dotted), and $\mathcal{R}$=0.7 (dot-dashed). Jets are corrected by CAFIX 5.0. (a) Jet finding efficiency vs. quark E_T. (b) Jet finding efficiency vs. quark η. (c) Jet energy scale linearity with quark E. (d) Reconstructed mass of the top quark from correct jet combinations. The area reflects the overall efficiencies.

tained from the measured energy (E_{meas}) as follows:

$$E_{meas} = R_{had}(E, \eta, \mathrm{RMS}_{\eta,\phi}) \times E_{true} \times (1 + C(E, \eta)) + U + N + Z, \qquad (3.8)$$

where R_{had} is the overall hadronic response (or hadronic scale), which depends on the energy, η, and the width (RMS) of the jet; C is a correction for the fraction of unclustered energy that the jet algorithm applied either incorrectly, left out, or pulled into the jet. There are three additive offsets, U for the energy from underlying events, N for the noise, and Z for the zero-suppressed portion of the jet energy. U is defined as the amount of uncorrected energy that enters the jet and originates from the spectator-part of the interaction (forward fragmentation). N is defined as the amount of uncorrected energy that enters the jet, and originates from electronic or uranium noise. Z is defined as the amount of the uncorrected jet energy lost because the calorimeter is zero suppressed (not read out if signal is within some window). The hadronic scale is derived using the "missing p_T projection fraction" technique (MPF), developed initially by CDF[64]. It uses the idea that any component of $\not{E}_T$ along the photon direction is due to mismeasurement of the jet energy in "photon"+jets events. (In fact, the events used are not truly photon + jets events, but correspond to events with large electromagnetic energy deposition in a restricted region of the calorimeter and jets balancing these "photons".) Parametrizing this MPF effect as a function of the η (needed mainly due to the imperfections in the ICD region), the jet width (due to the nonlinear correlation between jet width and the jet response in the detector), and the jet E_T, provides the overall correction factor for jet energy. As a final step, corrections are determined for energy lost outside of the cone of the jet (denoted as C, and algorithm dependent).

As already demonstrated in Figs. 3.5 (c) and (d), the CAFIX correction is not

entirely sufficient for retrieving the energy of the parton-level. This is not so crucial in searching for a signal, but its effect can even be minimized in the mass analysis through imposition of proper additional corrections.

3.2.4 Missing Transverse Energy

If a weakly interacting particle (such as a neutrino) were produced with high p_T, the negative of the vector sum of the momenta of the detected particles would have the same momentum vector as the undetected particle. Since the center of momentum frame of the partonic interactions is not generally at rest in the laboratory frame, only the transverse component of the missing particle's four-momentum can be reconstructed (this is referred as $\not{E}_T$ or missing transverse energy).

The calculation of $\not{E}_T$ is based upon energy deposition in the individual cells of the calorimeters. The vector $\not{E}_{Ti}$ is defined as follows:

$$\not{E}_{Tx} = -\sum_{cells\ i} E_{Txi} \tag{3.9}$$

$$\not{E}_{Ty} = -\sum_{cells\ i} E_{Tyi}. \tag{3.10}$$

The sums are over all cells in the calorimeter. The missing transverse energy, $\not{E}_T$ is the magnitude of this vector. Once CAFIX is applied, this missing transverse energy is recalculated and stored for subsequent analysis.

3.2.5 Electrons and Photons

Electrons and photons[65] are reconstructed as clusters of energy in the EM section of the calorimeter. The cluster algorithm used to find EM clusters is identical to the nearest neighbor algorithm, described in Sect. 3.2.2, but optimized

for EM objects. As in the case of the jet algorithm, energies in the towers of the EM section of the calorimeter are used as inputs to the clustering algorithm. An EM cluster is required to have at least 90% of its energy in the electromagnetic section of the calorimeter, and at least 40% of the energy must be contained in a single tower. The reconstruction program searches for a track from the central detector that points from the interaction vertex to the EM cluster within a window of $\Delta\eta=\pm0.1$, $\Delta\phi=\pm0.1$. If such a track is found, the cluster is identified as an electron, otherwise it is identified as a photon.

3.2.6   Muons

Muons are identified as tracks in the muon chambers[66]. Analogous to the reconstruction in the calorimeter, muon reconstruction begins with conversion of hits and time information into three-dimensional points, and then finding combinations of hits pointing towards the interaction vertex. Due to the presence of the magnetic field between layers, reconstruction is handled separately for segments before and after the field. At the end, muon tracks are obtained through a global fit[66] that combines tracks in the central detector and the interaction point.

Chapter 4

Data and Simulated $t\bar{t}$ Samples

This chapter describes the data and the simulated events of $t\bar{t}$ signal used in our analysis. In addition, we study various kinematic parameters that will be used to differentiate between signal and background.

4.1 Data Sample

4.1.1 Multi-jet Trigger

The data samples were selected by imposing an online-trigger and an offline-filter requirements, known, respectively, as the JET_MULTI trigger and filter[67, 68]. The JET_MULTI trigger requires the presence of at least four calorimeter trigger towers (0.2 in $\Delta\eta$ by 0.2 in $\Delta\phi$) with $E_T > 5$ GeV for RUN IA (data taken during 1992-1993, where the "run-numbers" range from 50000 to 70000), and at least three large tiles[69] (0.8 in $\Delta\eta$ and 1.6 in $\Delta\phi$) with $E_T > 15$ GeV for RUN IB (data taken during 1993-1995, where the run-numbers range from 70000 to 93115). During RUN IB, additional requirements were imposed of having at least four trigger towers with

Table 4.1: Initial criteria for data selection. Level 1 trigger requirements varied during the RUN 1A/1B periods, but the efficiency remained in the range of 95-98 %.

Generic Sample			
General Conditions	Sequential Requirements	Effective cross section (nb)	Efficiency for m_t=180 GeV/c^2
Level 1 trigger	JET_MULTI Three large tiles $E_T > 15$ GeV	0.4 ± 0.1 μb	~ 1.00
Level 2 filter	JET_MULTI Five R=0.3 L2 jets $|\eta| < 2.5, E_T > 10$ GeV	20 ± 5 nb	0.93
Five jets Five jets	$\mathcal{R}$=0.3 jet cones in DØRECO $|\eta| < 2.5, E_T > 10$ GeV	20 ± 5 nb	0.92
Bad data	BAD_RUN.RCP	-	-
Cleanup	(see next section)	-	-
Generic H_T	$H_T > 115$ GeV $\mathcal{R}$=0.5 jet cones in DØRECO	5.3 ± 1.3 nb	0.87
Search Sample			
No isolated e or μ Six jets	$E_T{}^e < 20$ GeV, $p_T{}^\mu < 15$ GeV/c $E_T > 10$ GeV CAFIX $|\eta| < 2$	4.7 ± 1.2 nb	0.64
$H_T > 150$ GeV	jet $|\eta| < 2.0$, $E_T > 15$ GeV	1.8 ± 0.5 nb	0.53

$E_T > 7$ GeV, in order to reduce the triggering rate (to fit into the bandwidth of the trigger system) at high luminosities ($\geq 10^{31}$ events cm^{-2} s^{-1}). No vetos were imposed on the Main Ring related triggers. The JET_MULTI filter required five $\mathcal{R}$=0.3 jets at Level 2, with these jets being restricted to $|\eta| < 2.5$ and $E_T > 10$ GeV. Again, during RUN IB, another condition was added requiring the sum of the E_T of jets (defined as H_T) to be greater than 110 GeV, in order to reduce the data rate at high luminosities. The effects of these changes on the acceptance for top-quark events were studied using Monte Carlo simulations and were found to be negligible[68]. Table 4.1 shows the list of trigger and filter requirements, with the calculated efficiencies for $t\bar{t}$ production for $m_t = 180$ GeV/c^2, used to define "generic" and "search" data samples (see next section).

4.1.2 "Generic" and "Search" Samples

In addition to imposing JET_MULTI trigger and filter requirements, a set of offline selection criteria were used to reduce the data sample to manageable size, without affecting acceptance for $t\bar{t}$ signal. First, the sum of the transverse jet energies for the 5 jets was required to be greater than 115 GeV, where the sum included jets only with $|\eta| < 2.5$. Also, the following requirements were imposed to eliminate events with spurious jets due to spray from the Main Ring and from any hot cells in the calorimeter[68, 70]:

- Bad Runs

  Runs identified as having problems with failure of detector components or the readout system were rejected. This reduced the data sample by $\sim$ 4 %.

- Main-Ring Veto

  We simply rejected any events that had the MICRO_BLANK or MRBS_LOSS flags set. This reduced the data sample by $\sim$ 13 %. Figures 4.1 (a) and

(b) show the missing E_T vs. ϕ of the jet, before and after the rejection of Main-Ring events, respectively. Our criterion removed the spurious cluster of jets in the region where the Main Ring pierces the DØ detector ($1.6 < \phi < 1.8$).

- Unphysical Events

  Events with more than 4 TeV energy in the calorimeter, and events with the scalar E_T sum (H_T) for cells in the calorimeter, ICD, and MG, with $H_T > 1$ TeV were also removed. These restrictions removed $\sim 2\%$ of all events.

- CH and Hot-Cell fractions

  Events that had any jet for which the CH fraction of the calorimetric energy was more than 70%, were rejected as spurious (due to noise in the calorimeter electronics or due to fluctuations from radioactivity in the uranium absorber). Events containing a hot-cell fraction (HCF) of more than 0.5 were rejected. The HCF routine identifies cells with the largest(cell 1) and second largest (cell 2) E_T in a jet, and defines the HCF as $E_T(\text{cell } 1)/E_T(\text{cell } 2)$. Together, these criteria removed $\sim 1\%$ of the remaining data.

- η_{rms}

  An energy-weighted RMS deviation of the jets in η:

$$\eta_{rms}^2 = \frac{\sum_{jets} \eta^2 E_T}{\sum_{jets} E_T} - \left(\frac{\sum_{jets} \eta E_T}{\sum_{jets} E_T}\right)^2 \qquad (4.1)$$

  was required to be greater than 0.1 in order to reject events due to noise from grounding failures that led to spurious signals in cells in the outermost ring ($|\eta|$=1.4) of the ECEM calorimeters. This removed $< 1\%$ of the data.

The above criteria defined our "generic" sample of data. Table 4.1 lists these selection criteria. In this chapter we will assume that the generic sample repre-
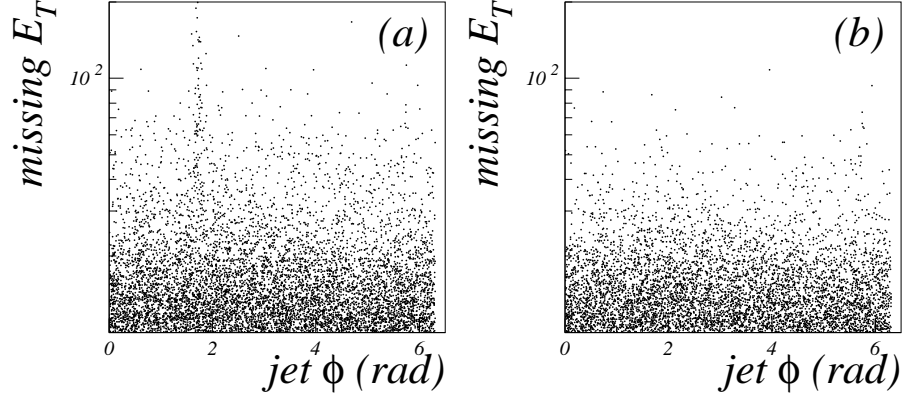
Figure 4.1: The effect of imposing the Main-Ring veto. A scatter plot of missing E_T vs ϕ for jets before (a), and after (b) imposing Main-Ring Veto.

sents the background. Because the fraction of the generic sample that is expected to originate from $t\bar{t}$ production is small (~ 0.001), such an assumption is quite reasonable.

The "search" sample is defined by imposing additional criteria. First, events containing isolated electrons or isolated muons ($E_T^e < 20$ GeV, $p_T^\mu < 15$ GeV/c) are removed. This is done in order to reject events that are already analyzed in other $t\bar{t}$ channels, such as in lepton+jets final states. Next, H_T is recalculated using CAFIX corrections to jets for $|\eta| < 2.0$ and $E_T > 15$ GeV, and the threshold for H_T is increased to 150 GeV, in order to further reduce the size of the data set, but still maintain high efficiency for the $t\bar{t}$ signal. Also 6 jets with E_T greater than 15 GeV are required. After these restrictions, the efficiency for ISAJET $t\bar{t}$ Monte Carlo for $m_t = 180$ GeV/c^2 production in the all-jets channel is 53 %.

4.1.3 Total Integrated Luminosity

Data obtained during the collider-run period of 1992-1995 (RUN I) are analyzed in this dissertation. Integrated luminosities were calculated using the routine GET_FILT_LUM[71], one of the standard DØ utilities. When GET_FILT_LUM returned an unphysical value of luminosity, then a nominal cross section of 7.5 nb was assumed for the JET_MULTI trigger which was corrected on the basis of the number of accepted events. After correction, RUN I data correspond to a total integrated luminosity of 95.3 events/pb with a fractional uncertainty of 5.4%.

4.2 Estimation of Background

The branching fraction of a b-quark into μ + anything is known as about $10.7 \pm 0.5\%$[11]. However, when all contributions from decays of b quarks and c quarks are considered in $t\bar{t} \rightarrow$ all-jets mode, approximately 50 % of events are expected to yield at least one muon. Therefore one expects to detect muon(s) originated from b-quark decay in $t\bar{t}$ events about 50% of the event (assuming a perfect detection efficiency). On the other hand, muons from QCD background process arise mainly from gluon splitting in $c\bar{c}$ or $b\bar{b}$ pairs which make up a small fraction of the background[72]. We will require the presence of at least one muon near a jet in every event (that is, "muon tagging") to enhance the $t\bar{t}$ signal, and will therefore also need a way to estimate the background in a given data sample. Because b-quarks from top decay have higher momenta than quarks from QCD events, we expect a factor of about ten improvement in signal to background ratio from requiring such muon tags within jets[72].

4.2.1 Muon Tagging of Jets

Procedures for muon tagging were defined initially after extensive Monte-Carlo studies of $t\bar{t}$ production in the lepton+jets final states[73]. The requirements used to select such muon tags are:

- The presence of a fully reconstructed muon track in the central quadrants of the muon system (CF). This restriction does not have much impact on the acceptance of b-quark jets from $t\bar{t}$ decay because these b-quarks tend to be produced mainly at central rapidities (See Fig. 1.3).

- The track must be flagged as a high-quality muon. This quality is based on the fit χ^2 to the track in both the bend and non-bend views of the muon system[74].

- The response of the calorimeter in the road defined by the track must be consistent with the passage of a minimum ionizing particle. Response is measured by energy deposited in the calorimeter towers nearest the track, and is required to be at least 1.5 times the energy expected for a minimum-ionizing particle.

- The muon energy must be greater than about 4 GeV in order to penetrate the material of the calorimeter and the iron toroids at 90° (η=0). Because the p_T spectrum of muons from π and K decays is softer than from heavy quarks[73], an overall $p_T > 4$ GeV/c cutoff is imposed to enhance the signal from heavy quarks.

- The muon must be reconstructed near a jet with $|\eta|$ <1.0 and $E_T > 10$ GeV, specifically, $\Delta R(\mu, jet) < 0.5$. This selects preferentially muons from heavy-quark semi-leptonic decays.
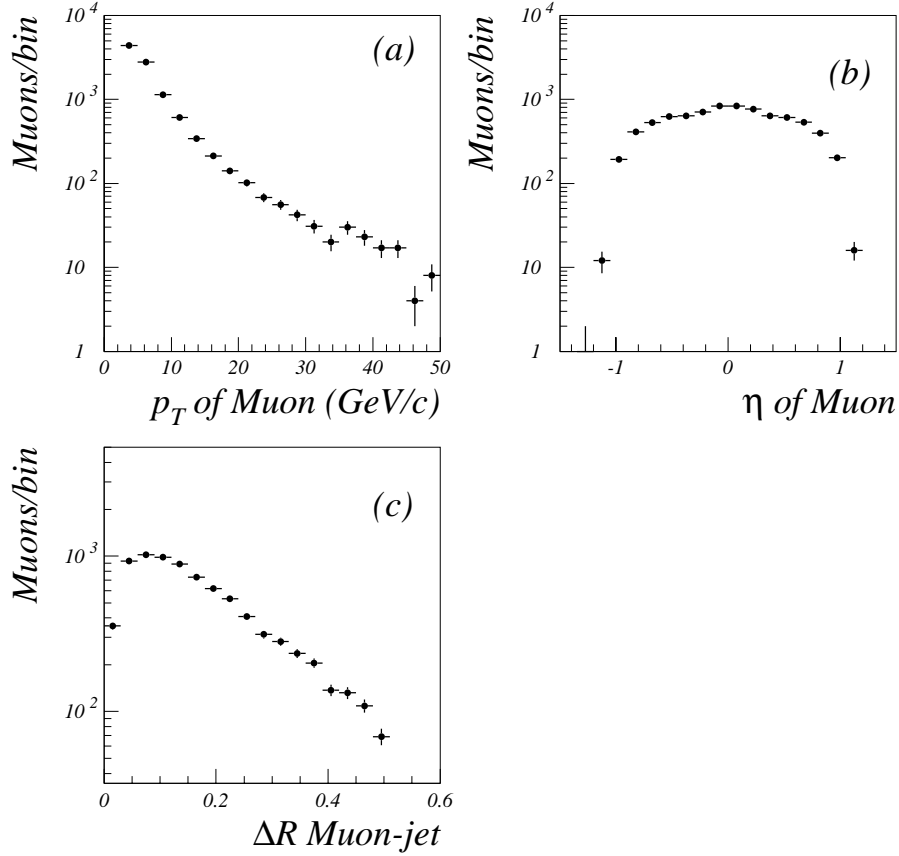
Figure 4.2: Characteristics of muons used for tagging jets in data of RUN I. (a) muon p_T , (b) muon η, and (c) the distance between the tag muon and the tagged jet in η-ϕ.

If a muon satisfies the above conditions, the jet associated with the muon is defined as a muon-tagged jet, and muon is called a tag. Figure 4.2 shows p_T and η distributions for tag muons, and the distance between the tag muon and the tagged jet in η-ϕ. The p_T spectrum appears to have an excess near ~ 40 GeV/c that can be attributed to contamination from a random overlap of $W \rightarrow \mu\nu$ decays and jets.
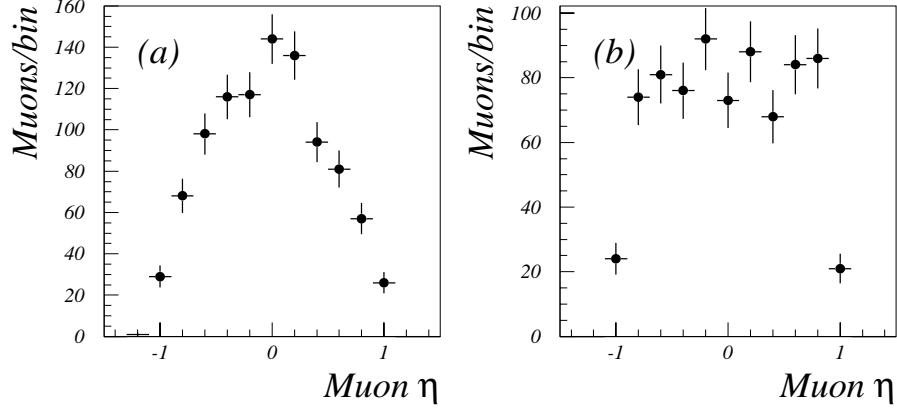
Figure 4.3: Pseudo rapidity distribution of tag muons before (a) and after (b) cleaning of muon chambers. The cleaning involved passing high currents through the wires[76].

4.2.2 Muon Tagging Rates

It is consistent with observation[72] that the probability of tagging general (background) events containing several jets is just the sum of the probabilities of tagging individual jets. Therefore, the muon tagging rate is defined in terms of probability per jet rather than per event. Initially[18, 72], the tagging rate was modeled only as a function of jet E_T[18, 72]. However, it was observed subsequently[75] that the η dependence of tag muons depended on run. This was traced to the fact that chambers experienced radiation damage during the run, and wires in some of the CF chambers had to be cleaned during the run[76]. Figure 4.3 shows the η distribution of muons for two different ranges of runs. Figure 4.3 (a) corresponds to the time before the cleaning and (b) to that after the cleaning (Runs $\geq$ 89000). It is clear that without taking this difference into account, one would lead to wrong estimates of tagging rates.

To address this problem, the tag-rate function for background $P_{tag}(E_T^{jet},\eta,\text{Run})$,

Table 4.2: Parameters from fits of Eq. (4.3) to the muon tag rates for different ranges of runs.

	p_1	p_2
Runs≤70000	-0.013 ± 0.002	0.0061 ± 0.0006
70000≤Runs<84000	-0.013 ± 0.001	0.0055 ± 0.0005
84000≤Runs<89000	-0.010 ± 0.002	0.0046 ± 0.0005
Runs≥89000	-0.010 ± 0.001	0.0045 ± 0.0005

was made a function of E_T^{jet}, η and the run-number, and assumed to factorize[75, 78]:

$$P_{tag}(E_T^{jet}, \eta, Run) = f(E_T^{jet}, Run) \cdot g(\eta, Run), \qquad (4.2)$$

where $f(E_T^{jet}, Run)$ is the probability that a jet has a muon tag in the very central region ($|\eta| < 0.3$), and $g(\eta, Run)$ is the measured chamber efficiency. Figures 4.4 (a), (b), (c), and (d) show $f(E_T^{jet}, Run)$ for different ranges of Run Numbers (respectively, for Runs < 70000, 70000 ≤ Runs < 84000, 84000 ≤ Runs < 89000, Runs ≥ 89000). Besides the differences in tag rates caused by the cleaning of wires, there were also changes in the gas mixtures used in the muon chambers between RUN 1A and RUN 1B, and changes in the high voltage settings implemented near Run 84000, which required the two additional separations of runs shown in Fig. 4.4[77]. Although statistics are poor, the change in the central tag-rate is substantial between RUN IA and RUN IB. The dependence of the muon tag rate on jet E_T in Fig. 4.4 is parametrized by a logarithmic form:

$$f(E_T^{jet}, Run) = p_1 + p_2 \log(E_T). \qquad (4.3)$$

The parameters p_1 and p_2 were obtained from fits to the different ranges of runs, and are listed in Table 4.2. The ratios of η distributions of tagged to untagged jets, representing an approximation to the relative chamber efficiencies $g(\eta, run)$ for the
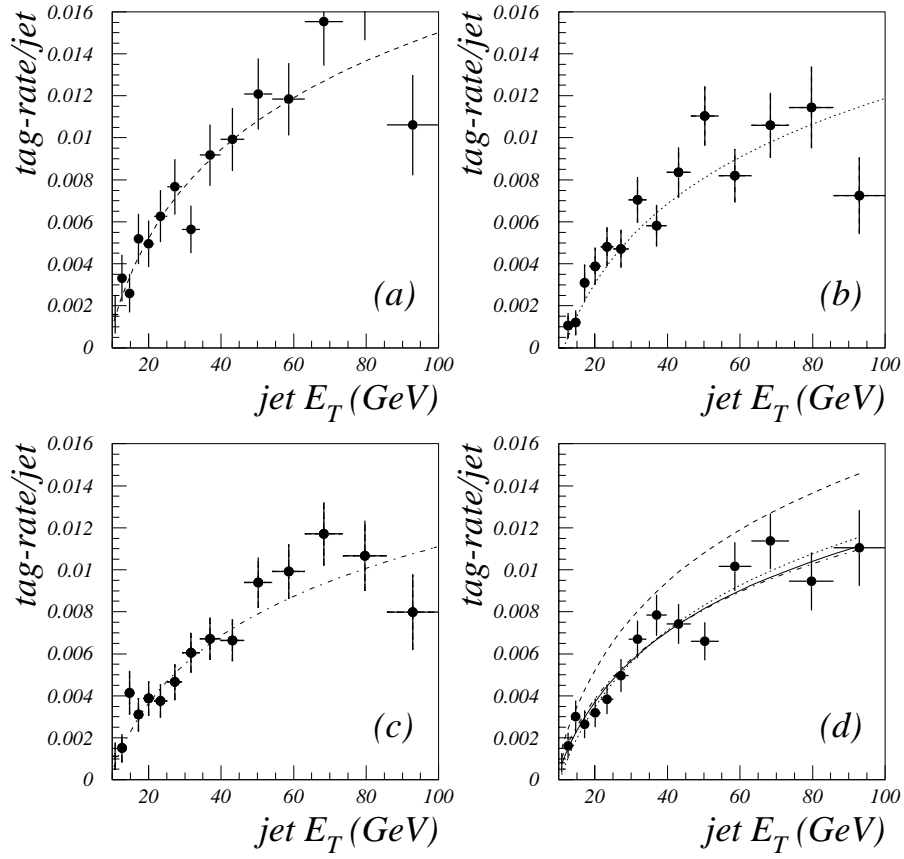
Figure 4.4: The muon tag rate (probability) for generic jets with $|\eta| < 0.3$, as a function of E_T^{jet}, for different ranges of runs. Curves represent the results of a fit to the data of (a) Runs < 70000, (b) $70000 \leq$ Runs < 84000, (c) $84000 \leq$ Runs < 89000, (d) Runs $\geq$ 89000. For comparison, the fits from (a), (b), and (c) are also shown in (d).

separations given in Fig. 4.4, are shown in Fig. 4.5. The crosses represent the ratio of observed tagged to non-tagged jets as a function of η. The curves are results of polynomial fit to the points. There seems to be a small asymmetry in the chamber efficiency in Fig. 4.5 (a) and (d). As a check, we ignored the asymmetry in Fig. 4.5 (a) and (d), and repeated fits to symmetric polynomials. That increased χ^2 by a factor of 4 in both cases.

The number of predicted tagged events from background can be written as:

$$N_{tag}^{pred} = \sum_{events} \sum_{jets} P_{tag}(E_T^{jet}, \eta, Run). \qquad (4.4)$$

Assuming that the heavy quark content is not significantly affected by the selection criteria, we expect the number of observed muon-tagged events to exceed the number predicted from background, and become more significant as we apply selection criteria to enrich the fraction of $t\bar{t}$ events. The reliability of estimating the number of tagged events expected from background can be evaluated by comparing the estimate to the results obtained in event samples that are dominated by background.

Figure 4.6 shows the distribution for the number of jets, E_T for the jets with fifth and sixth highest jet-E_T values, and the sum of all six jet transverse energies (H_T) for tagged 6-jet events (in black boxes). The predicted rate (in Eq. (4.4)) in all four distributions (in dashed lines) consistently matches the observed number of tagged events within a systematic uncertainty of $\sim 5\%$. It is hoped that, with an optimized combination of selection criteria, there may be regions of phase space where the contribution from $t\bar{t}$ production become especially significant, and produce a departure from expectations based on relation (4.4).
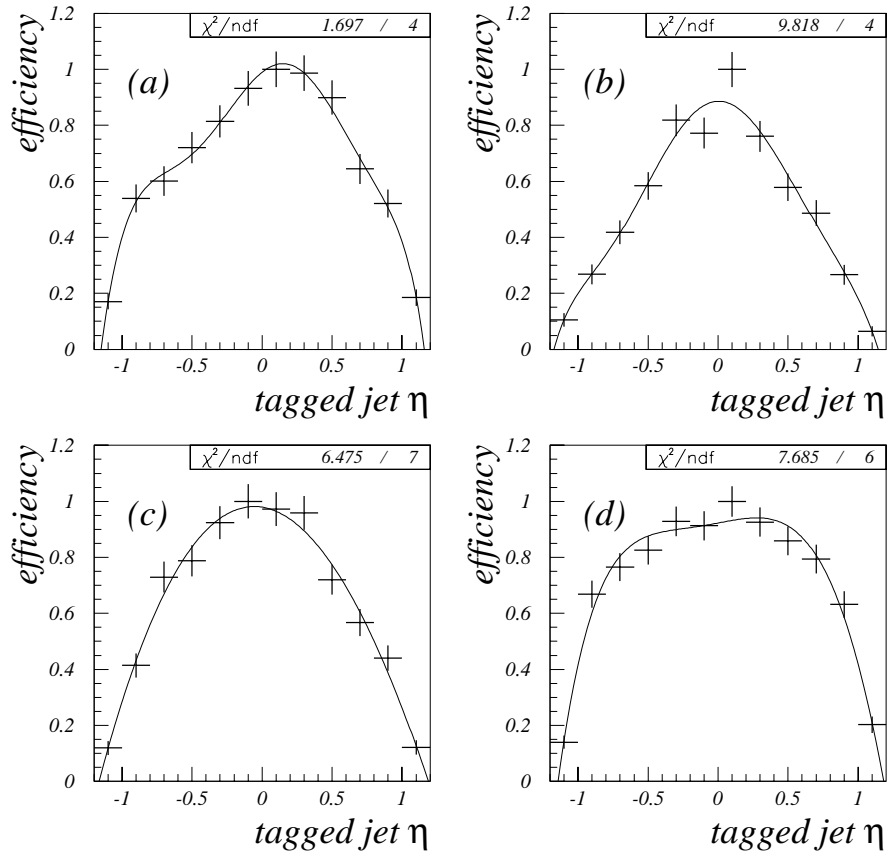
Figure 4.5: The measured muon chamber efficiencies (+) as a function of jet η and run number, and polynomial fits to those points for (a) Runs<70000, (b) 70000≤Runs<84000, (c) 84000≤Runs<89000, (d) Runs≥89000.
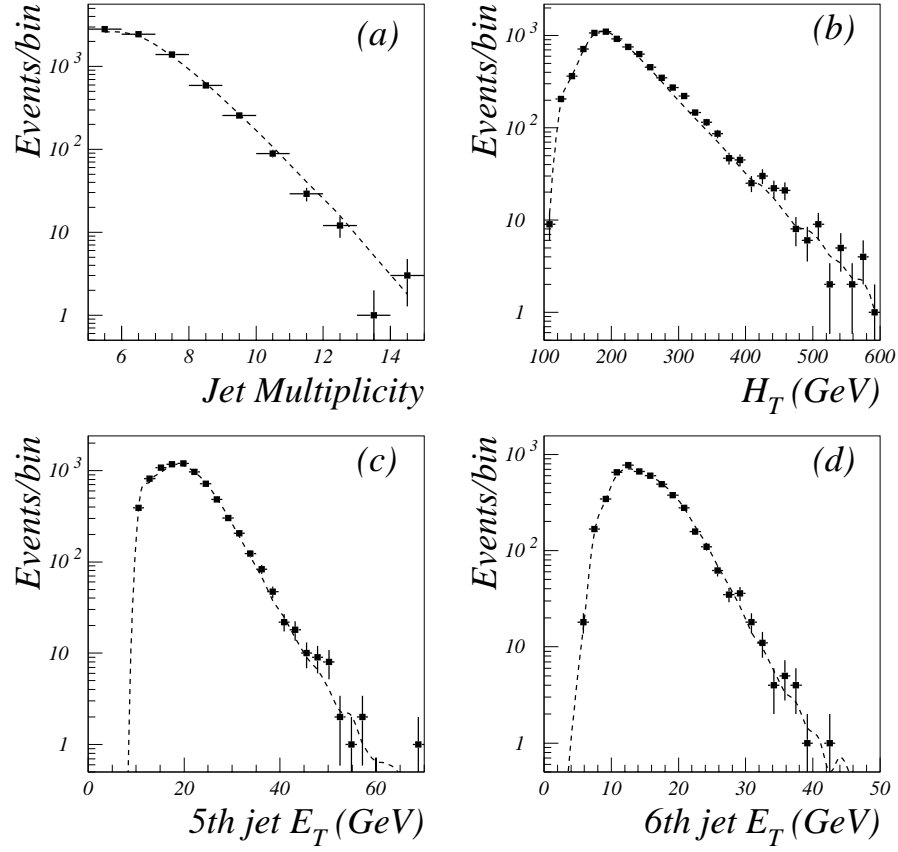
Figure 4.6: Comparisons of the number of muon tagged events expected from background (dashed) with muon tagged events (points with statistical errors), as a function of (a) jet multiplicity, (b) H_T, (c) E_T of jet with fifth highest E_T, and (d) E_T of jet with sixth highest E_T.

4.3   Simulations of Signal Events

Simulating $t\bar{t}$ events in the DØ detector is a difficult matter because of the ambiguities in the understanding of the production process and hadronization. In addition, there are still open questions regarding the modeling of the detector response, specially to low-energy particles, modeling of the underlying events, and modeling of the noise in the calorimeter. Different next-to-leading order (NLO) calculations appear to be consistent at the level of ≤ 20 %[13, 14, 15], giving confidence in the gross nature of $t\bar{t}$ production. In our analysis, we use primarily the ISAJET (VERSION 7.0) program to model and HERWIG (VERSION 5.7) to check such details. Although the two generators contain very different perspectives on hadronization, the efficiencies for detecting the $t\bar{t}$ signal, and the effect of imposition of a selection criteria needed to extract the mass of the top quark are in agreement, and the observed difference reflects systematic uncertainties of the analysis.

4.4   Kinematic Parameters

Extracting a signal from data dominated by background requires the study of kinematic parameters that can differentiate between features of the $t\bar{t}$ signal and the background. The most effective of these parameters will be studied in this section.

4.4.1   Aplanarity and Sphericity

The direction and shape of the momentum flow of jets in $t\bar{t}$ production is different from that in QCD background. These differences can be quantified by defining certain event-shape parameters[5, 79]. For each event, we can define the
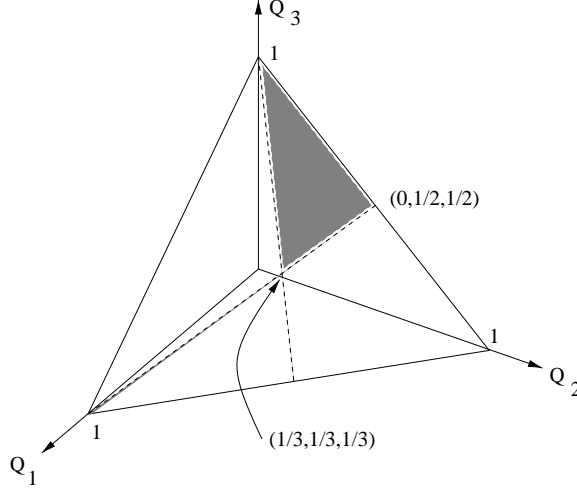
Figure 4.7: The allowed range of eigenvalues in the space spanned by Q_is.

normalized momentum tensor M_{ab}:

$$M_{ab} = \sum_i p_{(i)a} p_{(i)b} / \sum_i p_{(i)}^2 \tag{4.5}$$

where a and b run over the x, y, z components (index of the tensor), $p_{(i)}$ is the momentum of a given jet i, and i runs over the number of jets in an event. As is clear from its definition, M_{ab} is a symmetric matrix that is always diagonalizable, and has positive-definite eigenvalues (Q_1, Q_2, Q_3) satisfying the conditions[5, 79]:

$$Q_1 + Q_2 + Q_3 = 1, \qquad 0 \leq Q_1 \leq Q_2 \leq Q_3. \tag{4.6}$$

The equation $Q_1 + Q_2 + Q_3 = 1$ is a plane in a space spanned by Q_1, Q_2, and Q_3, and the inequality in the above equation restricts the range of each eigenvalue in a way that is clear from Fig. 4.7:

$$0 \leq Q_1 \leq \frac{1}{3} \tag{4.7}$$

$$0 \le Q_2 \le \frac{1}{2}$$
$$\frac{1}{3} \le Q_3 \le 1.$$

The magnitude of any Q_i represents the portion of momentum flow in the direction of axis i. Therefore, the event shape can be specified as follows:

- Linear : $Q_1 = Q_2 = 0$ and $Q_3 = 1$

- Planar : $Q_1 = 0$ and $Q_2 = Q_3 = 1/2$

- Spherical : $Q_1 = Q_2 = Q_3 = 1/3$.

The *aplanarity* ($\mathcal{A}$) and *sphericity* ($\mathcal{S}$) parameters that we will use are defined as follows:

$$\mathcal{A} \equiv \frac{3}{2} Q_1, \qquad \text{with } 0 \le \mathcal{A} \le 0.5 \tag{4.8}$$

$$\mathcal{S} \equiv \frac{3}{2} (Q_1 + Q_2), \quad \text{with } 0 \le \mathcal{S} \le 1.0. \tag{4.9}$$

Top-quark ($t\bar{t}$) events tend to have higher aplanarity and sphericity than background events. Although the actual values of the aplanarity and the sphericity are calculated in the center-of-momentum frame of the colliding partons, in terms of discrimination relative to QCD background, little difference was observed when $\mathcal{A}$ and $\mathcal{S}$ were calculated in the $\bar{p}p$ collision frame.

Figure 4.8 shows the distributions of these two parameters in three different ways. First, (a) and (b), show the distributions normalized to the production cross section, while (c) and (d) show both the distributions normalized to unity (densities). Finally (e) and (f) show the log of the probability ratios of the signal

to the background, $\log(R)$, with the ratio of integrated probabilities defined as:

$$R(x_{thr}) \equiv \frac{\displaystyle\int_{x_{thr}}^{x_{max}} \rho_{sig}\, dx\prime}{\displaystyle\int_{x_{thr}}^{x_{max}} \rho_{bkg}\, dx\prime}, \tag{4.10}$$

where $\rho(x)$ are the normalized densities for signal (sig) and background (bkg) for the parameter x, and x_{thr} and x_{max} are the thresholds and upper limits for the parameter x. The degree to which $\log(R)$ changes as x_{thr} increases, corresponds to the effectiveness of that parameter in discriminating against background events. Both $\mathcal{A}$ and $\mathcal{S}$ provide some degree of rejection of the background. The $\log(R)$ values change by a factor of ~ 10 within the allowed range of the parameters, and are not very sensitive to the specific value of the input top-quark mass. Aplanarity appears to provide slightly better rejection. Naturally, greater discrimination comes at the loss of signal, and a compromise has to be struck where to place any cutoffs.

4.4.2   Mass Sensitive Parameters

Any parameter that depends on the energy scale of jets is in principle also sensitive to the mass of the top quark. These "mass sensitive" parameters usually provide better discrimination against QCD background than parameters such as aplanarity and sphericity. This is because the characteristic energy scale (commonly called Q^2) of the QCD background is much lower than for the case of top quark production. That is, the QCD background is basically a 2 to 2 parton process that contains additional radiated gluons. Four mass-sensitive parameters will be defined and examined in this dissertation:

- H_T

  A sum of the transverse energies of jets in a given event characterizes the
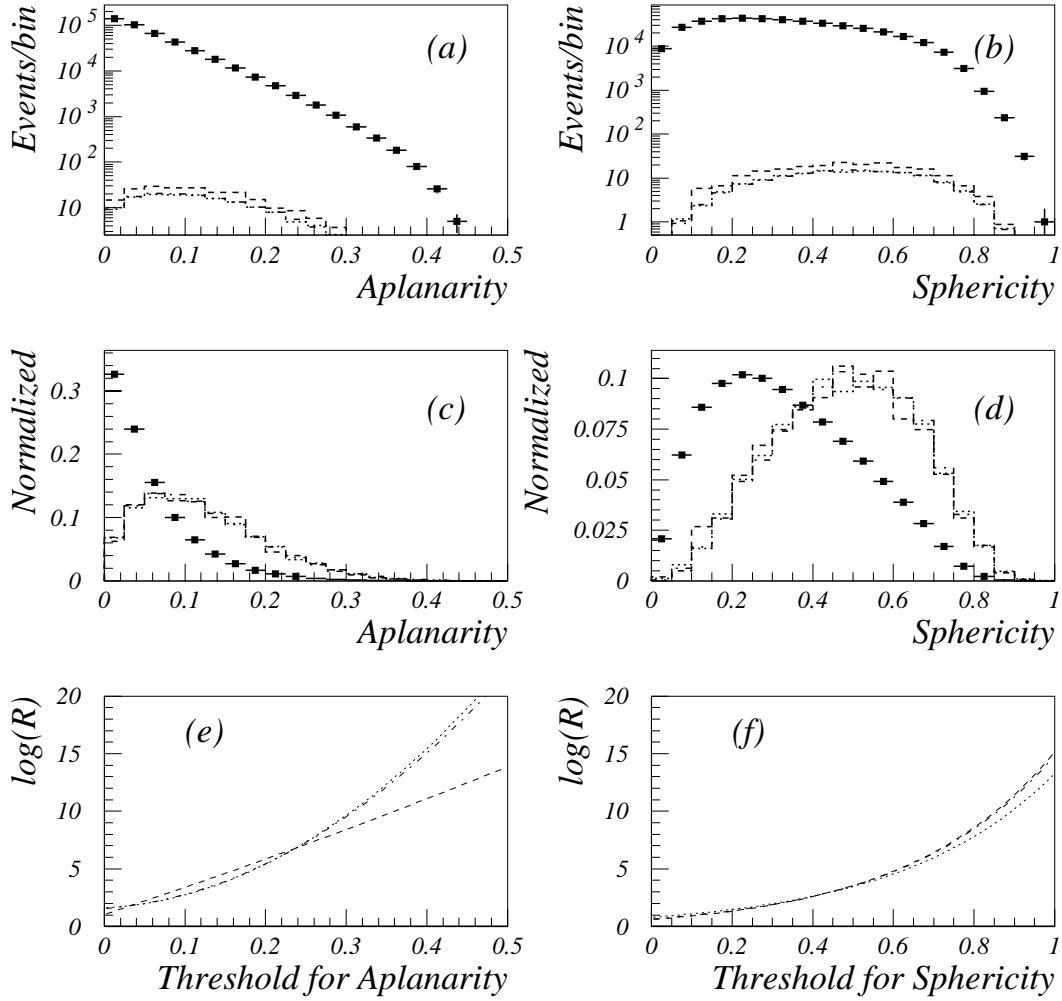
Figure 4.8: The aplanarity and sphericity distributions for data from RUN I (black boxes), and are from ISAJET $t\bar{t}$ events with a top-quark mass of, 140 (dashed), 160 (dotted), 180 GeV/c^2 (dot-dashed). (a) Aplanarity distributions normalized to their respective cross sections, (b) sphericity distributions normalized to their respective cross sections, (c) aplanarity distributions normalized to unity, (d) sphericity distributions normalized to unity, (e) log of the probability ratios for aplanarity, (f) log of the probability ratios for sphericity.

transverse energy flow, and is defined as before:

$$H_T \equiv \sum_j^{jets} E_{T(j)} \tag{4.11}$$

where $E_{T(j)}$ is the transverse energy of the jth jet. The average value of H_T has an relation with the mass of the top quark (m_t) as, $m_t \sim < H_T > /2$.

- H_T^{3j}

  H_T^{3j} is defined as[80]:

$$H_T^{3j} \equiv H_T - E_{T(1)} - E_{T(2)} \tag{4.12}$$

where $E_{T(1)}$ and $E_{T(2)}$ are the transverse energies of the first two leading jets (highest E_T). The reason for subtracting the E_T of the two leading jets is that QCD background consists of two hard-scattered partons with four soft gluons. Consequently, removing the two leading jets from the event should enhance the discrimination power. Figure 4.9 shows the distributions in H_T and H_T^{3j} in the same manner as given in the previous plot of $\mathcal{A}$ and $\mathcal{S}$. Both distributions are clearly sensitive to the mass of the top quark. The $\log R$ for H_T changes by a factor of ~ 30, which indicates that it is a stronger discriminator than $\mathcal{A}$ or $\mathcal{S}$. The $\log R$ for H_T^{3j} is even stronger than H_T, and changes by a factor of ~ 300 in its range. The distributions of H_T and H_T^{3j} for $t\bar{t}$ production are very strongly correlated with the top-quark mass, as is clear in the plots of $\log R$.

- N_{jets}^A

  Another way of looking at the transverse energy flow in an event is to relate the energy flow to the number of jets observed. Tkachov[81] has recently
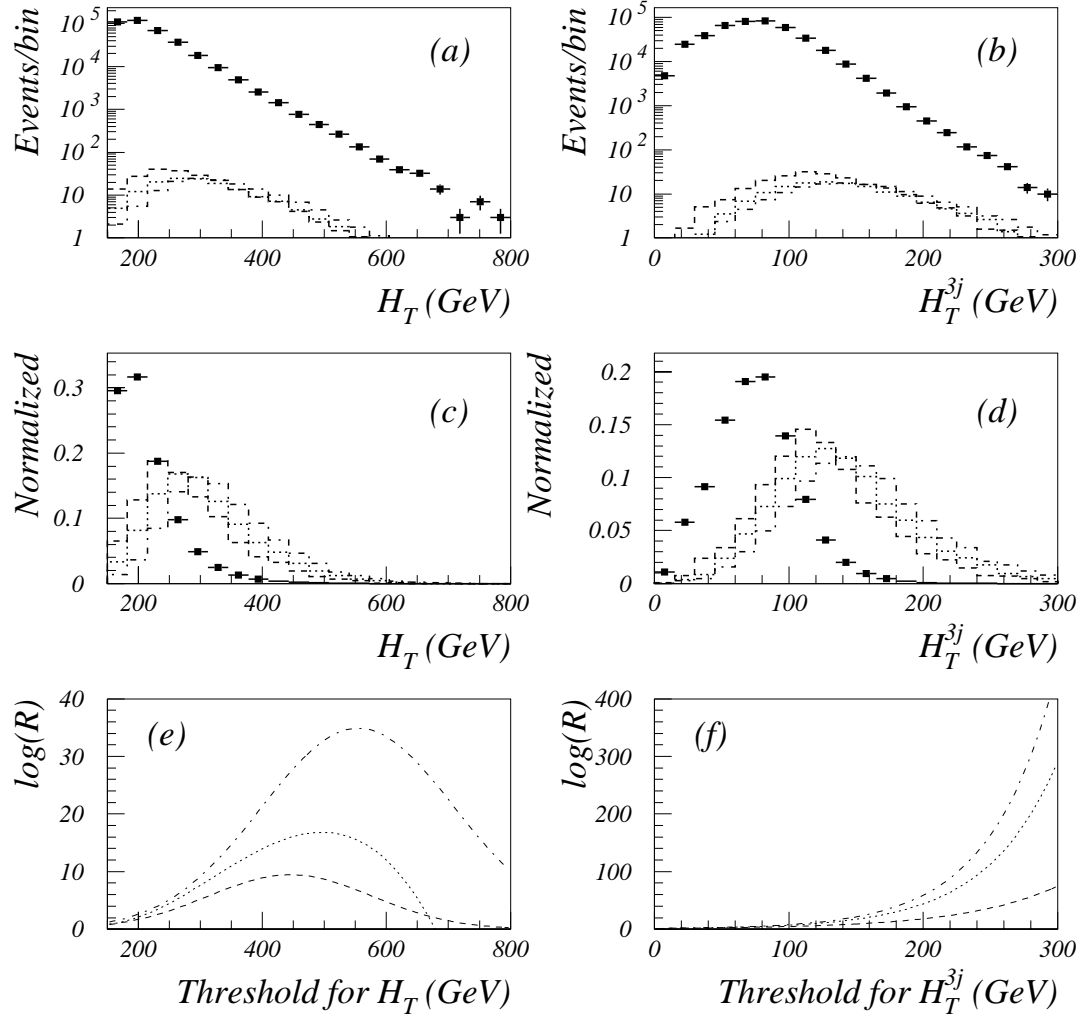
Figure 4.9: The H_T and H_T^{3j} distributions for data from RUN I (black boxes), and ISAJET $t\bar{t}$ events with a top-quark mass of 140 (dashed), 160 (dotted), 180 GeV/c^2 (dot-dashed). (a) H_T distributions normalized to their respective cross sections, (b) H_T^{3j} distributions normalized to their respective cross sections, (c) H_T distributions normalized to unity, (d) H_T^{3j} distributions normalized to unity, (e) log of the probability ratios for H_T, (f) log of the probability ratios for H_T^{3j}.

explored kinematic properties of events without defining their jets. Inspired by this work, an average jet-count parameter, N_{jets}^A, was developed by C. Stewart[70, 75]:

$$N_{jets}^A = \frac{\int E_T^{thr} N(E_T^{thr}) \, dE_T^{thr}}{\int E_T^{thr} \, dE_T^{thr}}, \tag{4.13}$$

where $N(E_T^{thr})$ is the number of jets in a given event for $|\eta| < 2.0$ and $E_T > E_T^{thr}$. Thus, this parameter is the number of jets averaged over a range of E_T thresholds (typically from 15 to 45 GeV), and weighted by the E_T threshold. This parameter corresponds to the number of jets, but is more sensitive to jets of higher E_T than just a jet count above some given threshold would be. The integration in the above equation is performed analytically using the following observation[75].

Suppose that $E_{T(low)}$ and $E_{T(high)}$ are the lower and upper limits of integration. Let n be the jet ranking order, where $E_{T(n)}$ is the closest in value to, but less than, $E_{T(high)}$. Then, for the n jets between $E_{T(low)}$ and $E_{T(high)}$, we can write (assuming no jets have E_T less than $E_{T(low)}$ and the number of jets is N):

$$
\begin{aligned}
&\int_{E_{T(low)}}^{E_{T(high)}} E_T^{thr} N(E_T^{thr}) dE_T^{thr} \tag{4.14} \\
=\ &\int_{E_{T(low)}}^{E_{T(1)}} E_T^{thr} N dE_T^{thr} + \int_{E_{T(1)}}^{E_{T(2)}} E_T^{thr}(N-1) dE_T^{thr} + ... \\
&+ \int_{E_{T(n)}}^{E_{T(high)}} E_T^{thr}(N-n) dE_T^{thr} \\
=\ &\frac{N}{2}(E_{T(1)}^2 - E_{T(low)}^2) + \frac{N-1}{2}(E_{T(2)}^2 - E_{T(1)}^2) + ... \\
&+ \frac{N-n}{2}(E_{T(high)}^2 - E_{T(n)}^2) \\
=\ &\frac{1}{2}\sum_{i=1}^{n} E_{T(i)}^2 + \frac{1}{2}(N-n)E_{T(high)}^2 - \frac{N}{2}E_{T(low)}^2.
\end{aligned}
$$

Therefore, N_{jets}^A can be calculated analytically using the following equation:

$$N_{jets}^A = \frac{\sum_{i=1}^{n} E_{T(i)}^2 + (N-n)E_{T(high)}^2 - NE_{T(low)}^2}{E_{T(high)}^2 - E_{T(low)}^2}. \qquad (4.15)$$

- E_T of the fifth jet

  The E_T of the fifth jet, $E_{T(5)}$, is also a useful parameter that discriminates QCD background from $t\bar{t}$ events.

Figure 4.10 shows distributions for N_{jets}^A and $E_{T(5)}$, again in the form used in Fig. 4.8. Both parameters can be used to reject background in an efficient way. In the case of N_{jets}^A, the signal to background ratio can be increased to $\sim 1/10$, as shown in Fig. 4.10 (a), but at a great cost of loss in statistics for the signal. Log R plots show that N_{jets}^A is typically about twice as powerful as $E_{T(5)}$ in discriminating against background. Both parameters are also strongly correlated with the top-quark mass.

4.4.3 Other Parameters

- $H_T/\sqrt{\hat{s}}$ and Centrality

  Two parameters that depend weakly on the mass of the top quark can be defined, as follows. Centrality, or $\mathcal{C}$:

$$\mathcal{C} = \sum_{j=1}^{N} E_{T(j)} / \sum_{j=1}^{N} E_{(j)} \qquad (4.16)$$

where the sums include all jets in the event with $|\eta| < 2.0$ and $E_T > 15$ GeV. This parameter is similar to H_T and also characterizes the transverse flow of energy in events, but is normalized in such a way that it depends weakly on
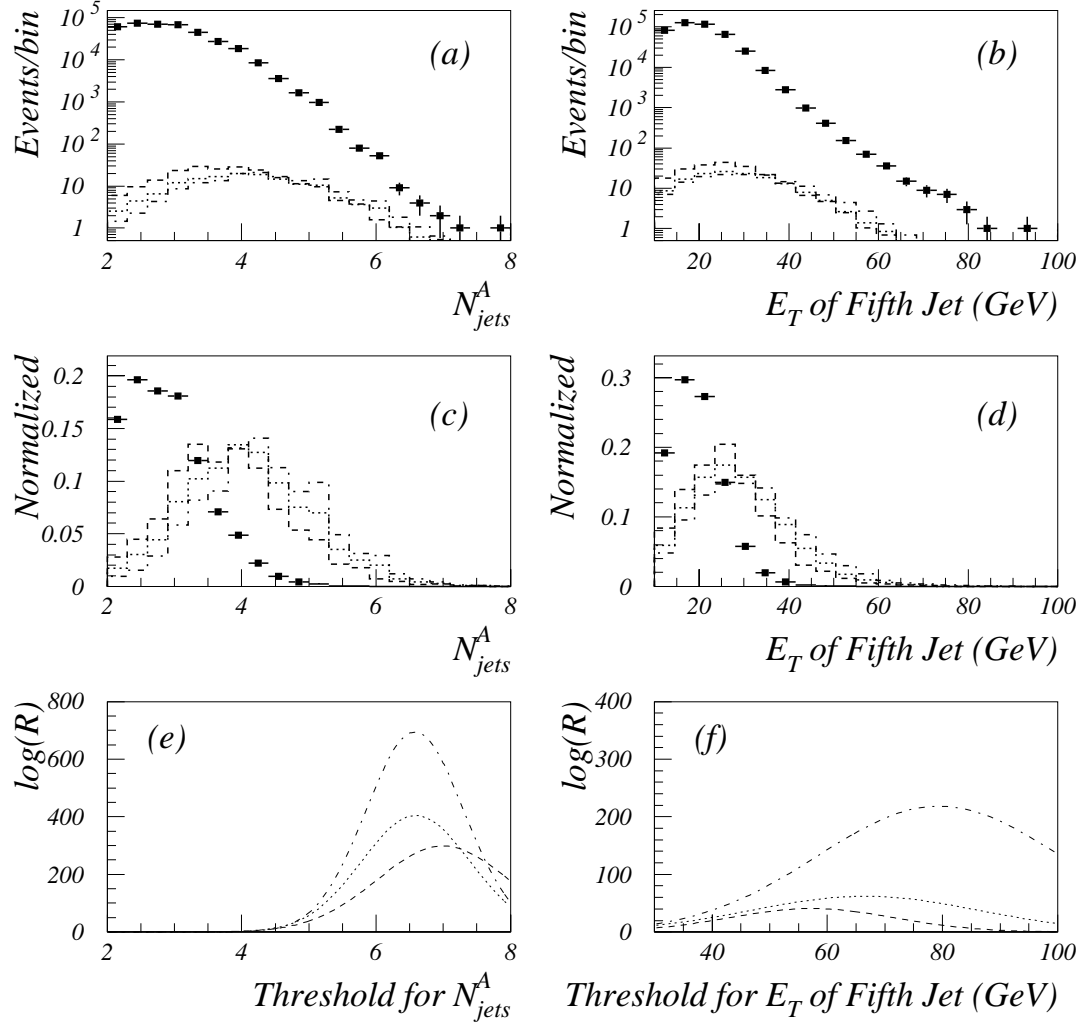
Figure 4.10: The N_{jets}^A and $E_{T(5)}$ distributions for data from RUN I (black boxes), and ISAJET $t\bar{t}$ events with a top-quark mass of 140 (dashed), 160 (dotted), 180 GeV/c² (dot-dashed). (a) N_{jets}^A distributions normalized to their respective cross sections, (b) $E_{T(5)}$ distributions normalized to their respective cross sections, (c) N_{jets}^A distributions normalized to unity, (d) $E_{T(5)}$ distributions normalized to unity, (e) log of the probability ratios for N_{jets}^A, (f) log of the probability ratios for $E_{T(5)}$.

the mass of the top quark. Second, we define a related quantity:

$$H_T / \sqrt{\hat{s}} \qquad (4.17)$$

where $\sqrt{\hat{s}}$ is the invariant mass of the colliding parton system. This is similar to the centrality.

Figure 4.11 shows distributions of $H_T / \sqrt{\hat{s}}$ and Centrality in the format of Fig. 4.8. Both parameters are weak in rejecting background, but are almost totally uncorrelated with the top-quark mass.

4.4.4 Correlations Between Parameters

Not surprisingly, the parameters we have introduced are correlated with each other. To quantify the degree of correlation, we introduce an experimental linear-correlation coefficient, r defined as[82]:

$$r \equiv \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\left[N \sum x_i^2 - (\sum x_i)^2 \right]^{1/2} \left[N \sum y_i^2 - (\sum y_i)^2 \right]^{1/2}}. \qquad (4.18)$$

The value of r ranges from 0, when there is no correlation, to ± 1, when there is complete correlation. Table 4.3 shows the correlations among the 8 parameters defined in the previous section. N_{jets}^A-$H_T/\sqrt{s}$ and $\mathcal{A}$-H_T are the least correlated pairs. Also, as expected, H_T^{3j}-$E_{T(5)}$ and H_T^{3j}-N_{jets}^A are very strongly correlated pairs. (The statistical uncertainties on the values of r are quite small.)

We will use H_T^{3j} as one of our choice of parameters because of its high rejection power (N_{jets}^A is dropped because of the spikes at every integral values in Fig. 4.10 (a)). Second, we take $\mathcal{C}$ as opposed to $H_T/\sqrt{s}$ because it provides less correlation with H_T^{3j}. We take $\mathcal{A}$ as opposed to $\mathcal{S}$ because it provides higher rejection power.
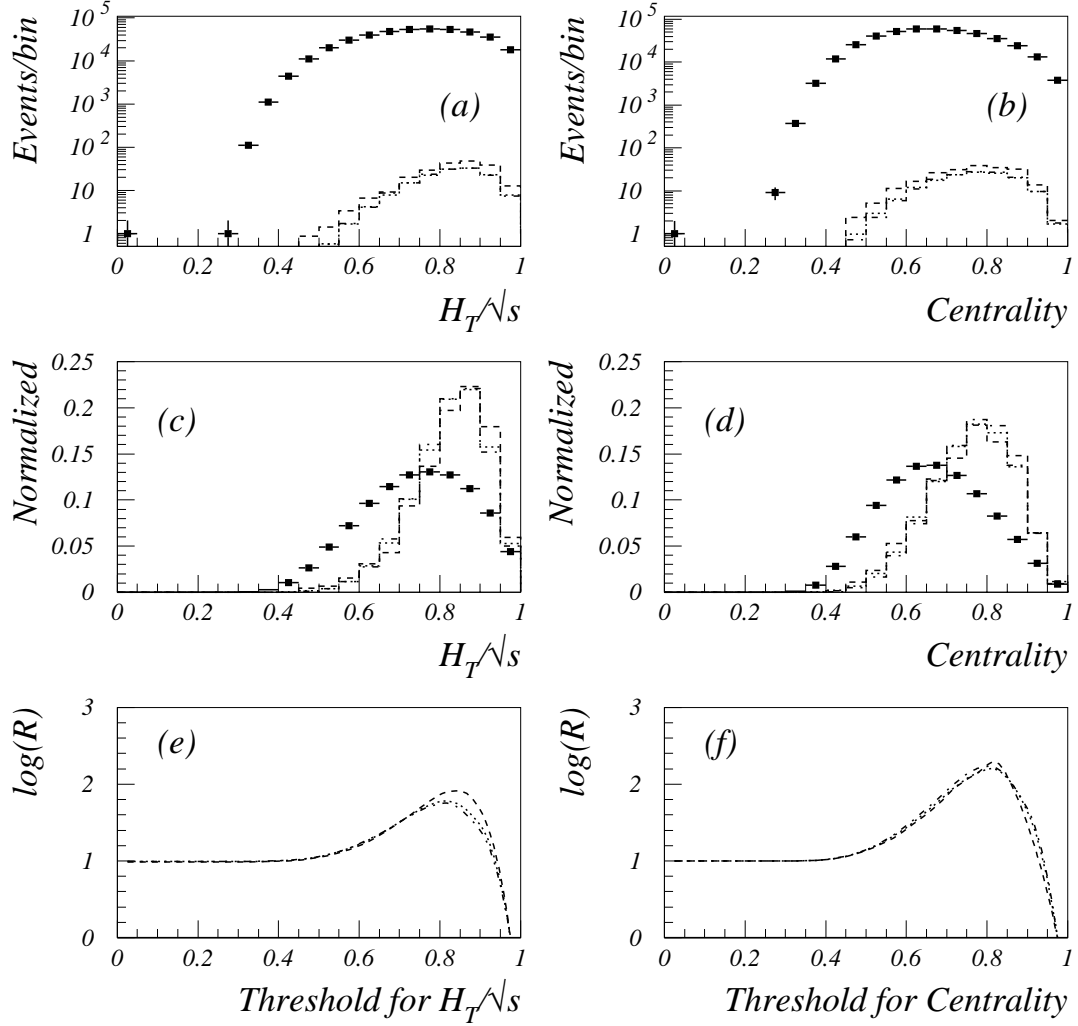
Figure 4.11: The $H_T/\sqrt{\hat{s}}$ and centrality distributions for data from RUN I (black boxes), and ISAJET $t\bar{t}$ events with a top-quark mass of 140 (dashed), 160 (dotted), 180 GeV/c^2 (dot-dashed). (a) $H_T/\sqrt{\hat{s}}$ distributions normalized to their respective cross sections, (b) centrality distributions normalized to their respective cross sections, (c) $H_T/\sqrt{\hat{s}}$ distributions normalized to unity, (d) centrality distributions normalized to unity, (e) log of the probability ratios for $H_T/\sqrt{\hat{s}}$, (f) log of the probability ratios for centrality.

Table 4.3: Correlations among kinematic parameters[70] for RUN I data.

	$\mathcal{A}$	$\mathcal{S}$	H_T	H_T^{3j}	N_{jets}^A	$E_{T(5)}$	$H_T/\sqrt{s}$	$\mathcal{C}$
$\mathcal{A}$	1	0.59	0.06	0.38	0.23	0.29	0.06	0.24
$\mathcal{S}$		1	0.06	0.24	0.18	0.19	0.11	0.39
H_T			1	0.69	0.74	0.28	−0.10	0.16
H_T^{3j}				1	0.80	0.61	−0.16	0.01
N_{jets}^A					1	0.42	0.14	0.03
$E_{T(5)}$						1	0.08	0.02
$H_T/\sqrt{s}$							1	0.37
$\mathcal{C}$								1

Finally we will use H_T and $E_{T(5)}$ in our analysis. The following chapter will describe how selection criteria on $\mathcal{A}$, $\mathcal{C}$, H_T^{3j}, $E_{T(5)}$ and H_T can be used to enhance the signal relative to background in the data.

Chapter 5

Search for $t\bar{t}$ Signal

In this chapter, we will focus on various techniques used to enhance and measure the contribution from $t\bar{t}$ production in the data.

5.1   Importance-Sampled Grid Search

Separating signal from background (or, more generally, event classification) by applying some set of restrictions has been a conventional way to enrich the signal in a given mixture of signal and background events. Let us assume that an event can be characterized by an n-tuple $X=(x_1,...,x_n)$ of parameters, (often referred to as a feature vector) and consider $(z_1,...,z_n)$ as a set of cutoffs on the x_i. In order to obtain an optimal set of cutoffs, a technique was developed to search for these cutoffs over a grid of points in a space spanned by $(x_1,...,x_n)$[83]. An efficient way to perform such a search is by using the distribution of the points $X=(x_1,...,x_n)$ from some appropriate subset of the data as a set of cutoffs. Because the distribution of data points corresponds to relevant physics processes, in this way one does not waste computer time in exploring uninteresting parts of phase

space. The set of such cutoffs forms a grid with some arbitrary spacing between neighboring grid lines. This can be termed an *importance-sampled* grid. The algorithm then scans all the points $X=(x_1,...,x_n)$ provided by both the signal and the background sample, and calculates the number of signal and background events satisfying the condition $x_1 < z_1$, $x_2 < z_2$, ..., $x_n < z_n$, for each $(z_1, z_2,...,z_n)$ set of cutoffs. The choice of cutoff values can now be decided based on what one wishes to optimize. In this dissertation, we will attempt to optimize the signal to background ratio in the data.

5.1.1 Training

To find the optimal selection criteria, we sampled the space of cutoff points using a combination of background and signal events. For background, we selected 7219 events at random (10 % of total) from the search sample (this comprises our "background training sample"). For the signal, we used ISAJET Monte Carlo events with an input top quark mass of 180 GeV/c² and assumed the total cross section of 4.7 pb. As described in the previous chapter, each of the events in the background sample has a well-defined probability of having one or more tagged jets, and that tagging probability is used to weight the background events. The signal sample consists of 8000 events, for which we assume a constant muon-tagging probability of 0.20 for each event, independent of kinematics. In order to have reasonable statistics, we do not, in fact, require muon tagging at this stage in either sample, but use the a priori tagging probabilities for weighting the events.

Based on the arguments presented in the previous chapter, the grid search was performed using the parameters H_T^{3j}, $\mathcal{A}$, $\mathcal{C}$, $E_{T(5)}$, and H_T. The values of the cutoffs (thresholds) were chosen from each of the above 8000 signal and 7219 background events for the importance-sampled grid in five-parameter space. Figure 5.1 (a)
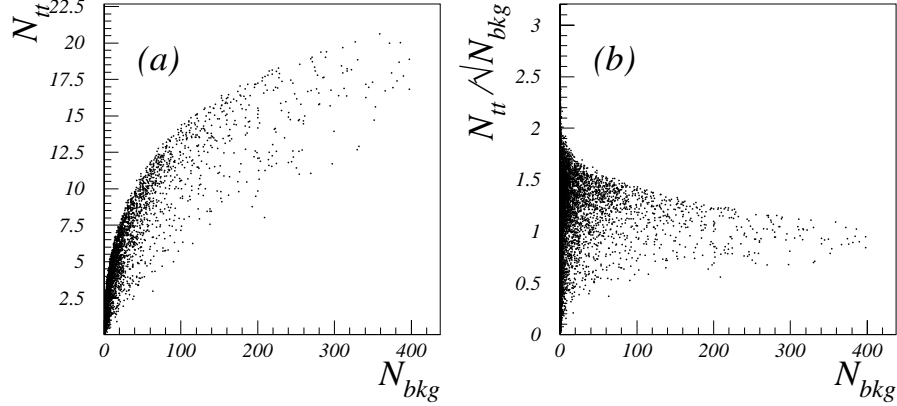
Figure 5.1: Results of the importance-sampled Grid Search in five-parameter space. Signal is modeled with ISAJET, using $m_t = 180$ GeV/c^2, and background is randomly selected from RUN I data. (a) Expected $t\bar{t}$ signal vs. expected background events for μ tagged jets, (b) significance $(N_{t\bar{t}}/\sqrt{N_{bkg}})$ as a function of the expected background events, for the full sample of $\approx$ 65,000 events ($\sim$ 72,000 - 7219 events).

shows the expected number of muon-tagged events $(N_{t\bar{t}})$ vs. the expected number of muon-tagged background events (N_{bkg}) for each set of 15,219 chosen thresholds, and Fig. 5.1 (b) shows the expected "significance" $(N_{t\bar{t}}/\sqrt{N_{bkg}})$ as a function of N_{bkg}. The number of signal and background events is scaled to the number expected to be in the search sample (excluding the 7219 events used for training) in Fig. 5.1. It is interesting to note that the grid search defines an upper edge in Fig. 5.1 (a) that corresponds to a family of parameter-threshold values that are optimal (optimal boundary). Our significance varies between 1 - 2, as can be seen in the upper points of Fig. 5.1 (b). To obtain the highest signal to background ratio for a given expected number of background events, we can choose any of the selection criteria defined by the points along this optimal boundary.

To check the need for using five (correlated) parameters, and to check any possible degradation of the signal to background ratio due to correlations between

Table 5.1: Threshold values for three points on the optimal boundary of the grid.

Selection Criteria	H_T^{3j} (GeV)	$E_{T(5)}$ (GeV)	$\mathcal{A}$	$\mathcal{C}$	H_T (GeV)
1	134	25	0.031	0.55	299
2	163	34	0.090	0.59	324
3	198	35	0.042	0.69	349

parameters, the grid search was performed again with only the three parameters, H_T^{3j}, $\mathcal{A}$, and $\mathcal{C}$. We found that using these three parameters yields a 5 % smaller signal to background ratio than using five parameters, for the case when one expects ~ 100 background events. Therefore, we will use the 5-parameter grid as our principal method in the search for signal.

5.1.2 Results

Assuming that the grid search extracts the best threshold values for any given number of expected background events, we chose three arbitrary points along the optimal boundary for our study. Figure 5.2 shows the chosen points. The three points are labeled as "set 1", "set 2", and "set 3". Table 5.1 lists the threshold values of H_T^{3j}, $\mathcal{A}$, $\mathcal{C}$, $E_{T(5)}$, and H_T that correspond to the three points.

In order to maintain statistical independence of the background-training and the search samples, we apply each of the three sets of thresholds to the search sample, excluding the data used for the training of the grid algorithm (7219 events). Table 5.2 lists the number of accepted events in RUN I (without muon tag), the number of predicted background events (with tag), the observed number of tagged events, and number of expected $t\bar{t}$ events with tags (for m_t=180 GeV/c^2, and assuming a total cross section of 4.7 pb), for the three grid points. The number of observed tagged events slightly exceeds the number of expected background events,
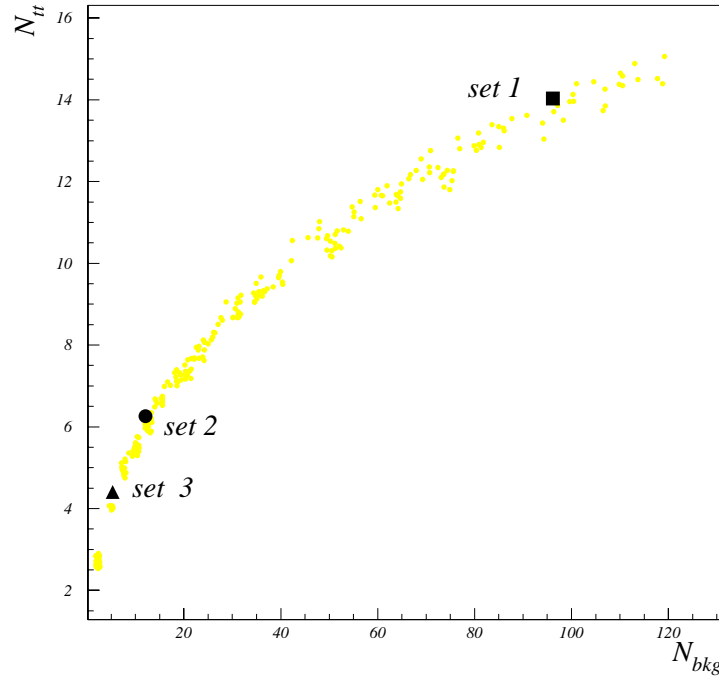
Figure 5.2: Region of the optimal boundary and three sets of cutoff points selected from the importance-sampled Grid Search of five-parameter space spanned by H_T^{3j}, $\mathcal{A}$, $\mathcal{C}$, $E_{T(5)}$, and H_T. Signal is modeled with ISAJET, using $m_t = 180$ GeV/c^2, and background is randomly selected data from RUN I. Three different symbols represent the three sets of cutoff to be used for extracting the size of the $t\bar{t}$ signal in the data.

Table 5.2: The number of untagged and tagged events in the search sample, the expected number of background events, and the $t\bar{t}$ signal expected from the grid search. The number of $t\bar{t}$ events also includes the contribution from other than all-jets channels which pass the selection criteria.

cut set	RUN I events (tag not required)	expected N_{bkg} (tagged)	observed (tagged)	expected $N_{t\bar{t}}$ (tagged)
1	2991	90.2	97	13.1
2	431	15.1	20	5.9
3	177	7.1	8	4.1

and this excess can be attributed to $t\bar{t}$ production. We will discuss the significance of the excess and the systematic uncertainties later in this chapter.

5.2   Analysis Based on the Covariance Matrix

The inverse of a covariance matrix C (a multi-dimensional generalization of the variance) can be used to reduce the dimensionality of an event-classification problem[84]. This reduction has two advantages: 1) it simplifies the optimization of selection criteria because there are fewer parameters to deal with, and 2) it tends to increase the efficiency with which events can be selected because it exploits linear correlations among the parameters[84]. The quadratic form $(x\text{-}\langle x\rangle)^T C^{-1}(x\text{-}\langle x\rangle)$ can be used as a tool for discriminating some specific class of n-tuple X from any other (where X refers to the same type of n-tuple we introduced in the beginning of our discussion of the grid search). It was shown that[84] the optimal way to separate any two overlapping distributions is to rely on a function f that is based on a common covariance matrix C that provides means $\langle X\rangle_B$ and $\langle X\rangle_S$:

$$f = \frac{1}{2}\left(\chi_B^2 - \chi_S^2\right),\tag{5.1}$$

where

$$\chi_B^2 \equiv (X - \langle X \rangle_B)^T C^{-1} (X - \langle X \rangle_B), \tag{5.2}$$

$$\chi_S^2 \equiv (X - \langle X \rangle_S)^T C^{-1} (X - \langle X \rangle_S), \tag{5.3}$$

$$C_{(i,j)} = \frac{1}{N} \sum_{k=1}^{N} (x_i^{(k)} - \langle x_i \rangle)(x_j^{(k)} - \langle x_j \rangle), \tag{5.4}$$

N is the number of events used in the construction of the covariance matrix, and the function, f is called the *Fisher's linear discriminant function*. It is certainly true that the quantities χ_B^2 and χ_S^2 are not useful if the parameter x_i has non-Gaussian nature in its distribution because that could mismodel C and consequently χ^2. For the application to our analysis, two covariance matrices, C_S and C_B, separately for the signal and the background were constructed as follows[84].

$$C_{S(i,j)} = \frac{1}{M} \sum_{k=1}^{M} (x_i^{(k)} - \langle x_i \rangle)(x_j^{(k)} - \langle x_j \rangle) \tag{5.5}$$

$$C_{B(i,j)} = \frac{1}{N} \sum_{k=1}^{N} (y_i^{(k)} - \langle y_i \rangle)(y_j^{(k)} - \langle y_j \rangle)$$

where the summation runs over events in each sample (M events for signal and N events for background), i and j represent elements of the covariance matrices, and x and y are the parameters representing signal and background sample, respectively. Using Eq. (5.5), χ_B^2 and χ_S^2 are redefined with their corresponding covariance matrices. We will use the Fisher's discriminant f to attempt to enhance the signal events from the $t\bar{t}$ production relative to background.

Before constructing covariance matrices, we first make the distributions more Gaussian by performing the following transformations[85]:

$$x_i \rightarrow x_i^p \tag{5.6}$$

$$x_i^p \to \frac{x_i^p - \langle x_i^p \rangle}{\sigma}$$

where p is a real number that is chosen in such a way that the distribution of x_i^p has no third moment (i.e., $\langle (x_i^p - \langle x_i^p \rangle)^3 \rangle = 0$), and σ is the RMS width of the distribution x_i^p. Figure 5.3 shows the results of such symmetrizations. To compare the changes in shapes, scales on the horizontal and the vertical axes are chosen arbitrarily (differently for distributions before and after applying the transformation). In Fig. 5.3 (a), we compare the H_T^{3j} distribution from the search sample before and after the transformation. The shaded histogram is the initial H_T^{3j} that is asymmetric with a long tail at its higher end. The transformed histogram is clearly non-Gaussian, but far less so than the input spectrum. The symmetrization is more effective for the aplanarity distribution of the search sample, as is shown in Fig. 5.3 (b). Figure. 5.3 (c) and (d) demonstrate the fact that the transformation also makes the distributions in H_T and C for the $t\bar{t}$ sample (ISAJET $m_t = 180$ GeV/c^2) closer to Gaussian in form.

5.2.1 Training

In order to determine the two covariance matrices C_S and C_B, we used the identical training sets of events that were used in the importance-sampled grid search ($t\bar{t}$ Monte Carlo for C_S, and background events for C_B). We did not require muon tagging in either sample, and in fact, ignored the tag probability in the construction of the covariance matrices. The same five parameters, H_T^{3j}, $\mathcal{A}$, $\mathcal{C}$, $E_{T(5)}$, and H_T, were used again in the analysis. Here, the presence of the correlation among parameters is less serious in the analysis, because the covariance matrix technique already takes proper account of any linear correlations.

Figure 5.4 shows the correlation between χ_B^2 and χ_S^2 for both background
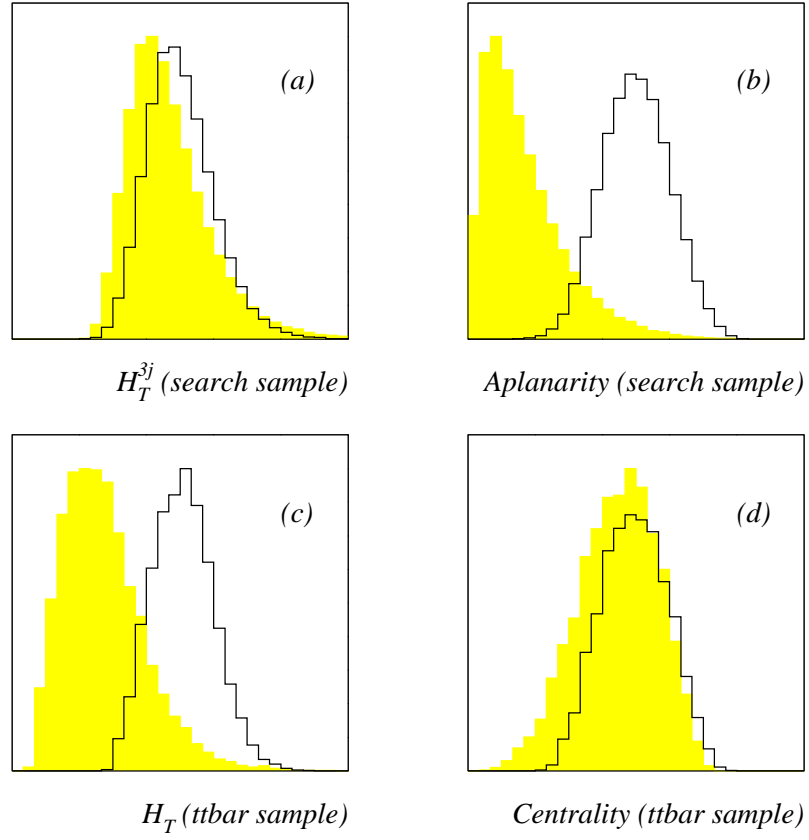
Figure 5.3: The result of symmetrizing kinematic parameters. Shaded histograms represent the original distributions, and the normal histograms represent the transformed distributions. Scales on axes are arbitrary, and chosen only to compare the shapes of the distributions. ISAJET $m_t = 180$ GeV/c^2 is used for $t\bar{t}$ sample. (a) H_T^{3j} distributions from the search sample, (b) $\mathcal{A}$ distribution from the search sample, (c) H_T distribution for $t\bar{t}$ sample, (d) $\mathcal{C}$ distribution for $t\bar{t}$ sample.

(shaded circles) and signal samples (crossed bars, representing ISAJET $t\bar{t}$ events at m_t of 180 GeV/c^2). We see a clear difference in the population of background and signal events in the (χ^2_B, χ^2_S) space, demonstrating the possible utility of the covariance matrix technique for distinguishing signal and background. Figures 5.5 (a) and (b) show the distributions for χ^2_B and χ^2_S, respectively, using background and $t\bar{t}$ signal samples. The signal is normalized to the number of events expected in the search sample after imposing muon tagging (and excluding the 7219 training events). As expected, the background sample has smaller χ^2_B values than the signal sample, and selecting χ^2_B can increase the signal to background ratio to $\sim 1/5$. The signal sample also has smaller χ^2_S values than the background (Fig. 5.5 (b)), but χ^2_S is not as strong a discriminator as χ^2_B. Figure 5.5 (c) shows the distributions in the Fisher's discriminant function for the $t\bar{t}$ signal (shaded histograms) and for the background (black circles). Using Fisher's discriminant function, one can reach a signal to background ratio of $\sim 1/2$ in the region where one expects to observe ≤ 10 events in the search sample. We will use Fisher's discriminant function as selection criteria for the search in the next section.

5.2.2 Results

We used an arbitrary set of integral values between 2 and 8 as thresholds for the Fisher's discriminant function, and simply counted the number of events that passed these thresholds, both for background and $t\bar{t}$ Monte Carlo. When applying the thresholds, we again excluded the data used for the training of the covariance matrices, in order to ensure the statistical independence of the training samples and the events used in the final analysis. Table 5.3 lists the threshold values, the number of events observed in RUN I for those thresholds (without tag), the number of expected background events (with muon tags), the number of observed tagged
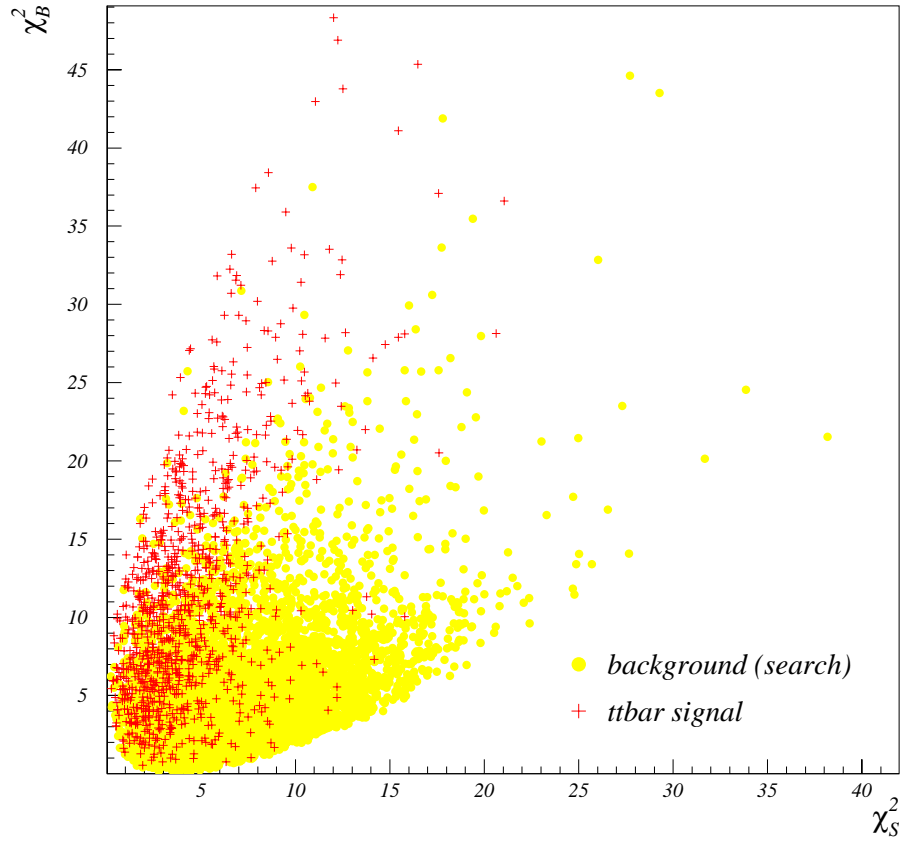
Figure 5.4: Correlation between χ_B^2 and χ_S^2 for the background-training sample ($\bullet$) and for the $t\bar{t}$ signal ($+$). The background training sample corresponds to 10 % of the search sample, and the signal sample is obtained from ISAJET Monte Carlo for $m_t = 180$ GeV/c^2.
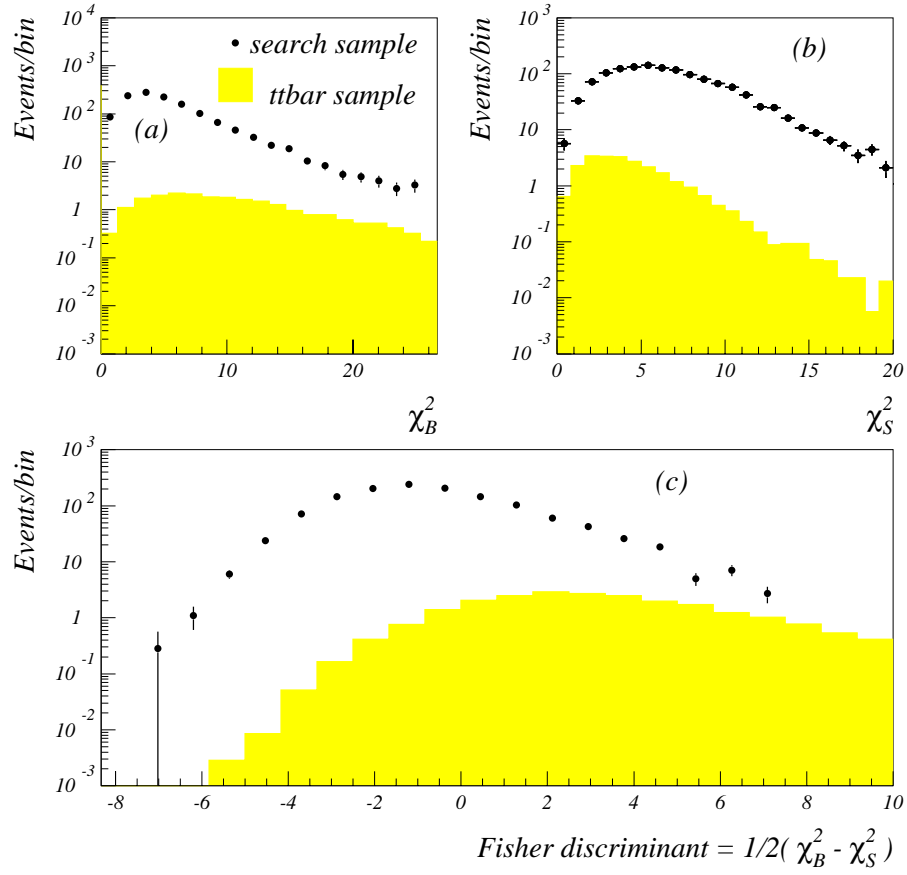
Figure 5.5: Distributions in χ^2_B, χ^2_S, and Fisher's discriminant function. Events are weighted to correspond to the expected numbers, after muon tagging. Shaded histograms refer to $t\bar{t}$ signal, and the points to background (search sample). (a) Distributions in χ^2_B, (b) distributions in χ^2_S, (c) distributions in the Fisher's discriminant function.
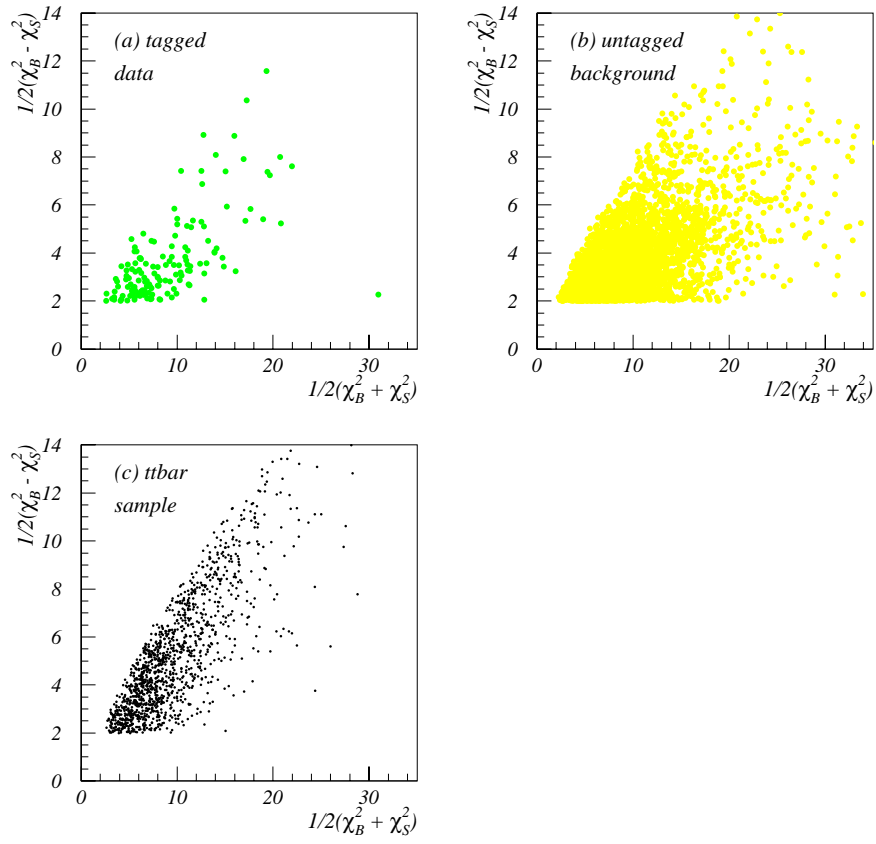
Figure 5.6: Scatter plots for data, background (untagged search sample) and $t\bar{t}$ signal of $\frac{1}{2}(\chi_B^2 - \chi_S^2)$ vs. $\frac{1}{2}(\chi_B^2 + \chi_S^2)$.

Table 5.3: The f dependence of the number of untagged events of RUN I, the expected tagged background, the observed number of tagged events, and tagged $t\bar{t}$, based on the covariance matrix techinique. (The number of $t\bar{t}$ events includes a small contribution from other than all-jets channels that pass the selection criteria.)

set	$f >$	events in RUN I (tag not required)	expected N_{bkg} (muon tagged)	observed (tagged)	expected $N_{t\bar{t}}$ (tagged)
a	2	4379	134.1	149	14.6
b	3	2279	74.6	81	11.4
c	4	1196	41.3	43	8.6
d	5	649	23.6	28	6.3
e	6	357	13.3	15	4.4
f	7	212	8.2	14	3.1
g	8	128	5.0	7	2.1

events, and number of tagged $t\bar{t}$ events expected for m_t=180 GeV/c^2, assuming a total cross section for $t\bar{t}$ production of 4.7 pb. It should be recognized that the number of observed events for increasing f-thresholds are not statistically independent, but are, in fact, subsets of one another. There is a small excess that can be attributed to signal above background, that will be discussed later in this chapter.

Figure 5.6 displays scatter plots of $\frac{1}{2}(\chi_B^2 - \chi_S^2)$ against $\frac{1}{2}(\chi_B^2 + \chi_S^2)$ for tagged data for RUN I, background (search sample), and $t\bar{t}$ signal, only for events with $f > 2$. The data in Fig. 5.6 (a) suggest the presence of a mixture of background, as given in Fig. 5.6 (b) and $t\bar{t}$ signal, as given in Fig. 5.6 (c). We will investigate in Appendix A the effect on kinematics of selecting different cutoffs for Fisher's variable.

5.3 Artificial Neural Networks Analysis

Artificial neural networks[86, 87] constitute a powerful nonlinear extension of conventional methods of multi-dimensional data analysis. The goal is to have a set of i outputs $F_i = F_i(x_1, x_2, ...x_n)$ of networks that provide different values for signal and for background. The goal is therefore identical to that of the grid search and of the approach using the covariance matrix. The word "network" refers to the weighted connections among internal parameters in F_i (commonly called nodes), and "feed-forward" means that the feature parameters (x_i) are used in only one direction, namely forward. The following form is often chosen for the F_i:

$$F_i(x_1, x_2, ..., x_n) = g \left[\frac{1}{T} \sum_j \omega_{i,j} g \left(\frac{1}{T} \sum_k \omega_{j,k} x_k + \theta_j \right) + \theta_i \right] \qquad (5.7)$$

which corresponds to the configuration of Fig. 5.7 (a). The "weights" $\omega_{i,j}$ and $\omega_{j,k}$ are parameters to be fitted using some input distribution. (Note that $\omega_{i,j}$ and $\omega_{j,k}$ are independent parameters. $\omega_{j,k}$ are weights connecting input to hidden nodes, and $\omega_{i,j}$ are weights connecting hidden to output nodes.) The functions in Eq. (5.7) are functions of their entire brackets. They are non-linear ("transfer" or "activation") functions, and typically of the form that we use[86]:

$$g(x) = \frac{1}{2}[1 + \tanh x]. \qquad (5.8)$$

This construction is, in fact, based on the structure of biological neurons[86], and that is why this technique is called "artificial neural" network analysis. The term θ_i is a threshold, corresponding to the membrane potential in a biological neuron, and is a parameter that also must be obtained from a fit. The parameter T (called "temperature") controls the steepness of the function g, as sketched in Fig. 5.7
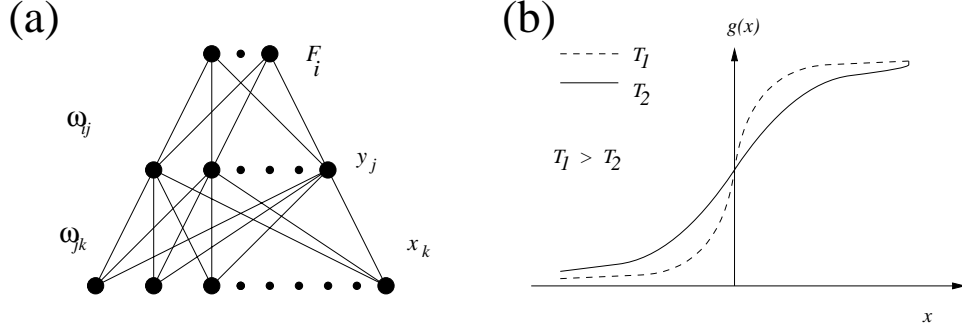
Figure 5.7: Neural basics. (a) A one hidden node layer feed-forward network architecture, (b) an example of non-linear neuron activation function.

(b). Depending on the values of the weights and thresholds, the contributions from each x_i to each output F_i will be different. The presence of the form $g[g(x)]$ (the "hidden layer"* in Fig. 5.7 (a) and the presence of $\omega_{i,j}$) in Eq. (5.7) provides for non-linear modeling of the input.

Practically, the goal is now to reduce the "error" between the desired response and the network's actual response. A commonly used error function is the mean square error E, averaged over the sample:

$$E = \frac{1}{2N_p} \sum_{p=1}^{N_p} \sum_{i=1}^{N} (F_i^{(p)} - t_i^{(p)})^2 \tag{5.9}$$

where t_i is the desired target value (the numerical choice for a particular class of events) for F_i, N_p is the number of patterns (number of events in the sample), and N is the number of network outputs. By performing the minimization of the function E, one can obtain the F_i that discriminate signal from background.

Many algorithms for the minimization (commonly called learning or training) of

---

*Input feature parameters (x_k) are mapped to another set of internal parameters, $y_j = g(\sum_k \omega_{j,k} x_k + \theta_j)$, as in Eq. (5.7). This set of parameters y_j is commonly called the "hidden layer" that ultimately determines the network outputs (F_i). Also, the necessity of only one hidden layer for a given event classification problem is guaranteed by a mathematical theorem[87].

the error function exist[88]. Back propagation is the most widely used learning algorithm since it is simple to implement and it often outperforms other methods[88]. It contains a gradient descent minimization method[88] that amounts to updating the weights and thresholds according to the back-propagation learning rule[88]. (The back-propagation learning rule will be explained in Appendix B.) During the minimization, weights and thresholds are updated as:

$$\omega_{t+1} = \omega_t + \Delta\omega_t \tag{5.10}$$

where

$$\Delta\omega_t = -\eta\frac{\partial E_t}{\partial\omega}. \tag{5.11}$$

Here ω refers to the vector of weights and thresholds used in a given network, and t and $t+1$ refer to the previous and the current updating, respectively. The quantity η is called the learning-rate parameter that controls the rate of the learning[87]. A typical values of the learning-rate are in the range [0.01,1]. A "momentum" term is often added[87]:

$$\Delta\omega_{t+1} = -\eta\frac{\partial E_t}{\partial\omega} + \alpha\Delta\omega_t. \tag{5.12}$$

where α is called the momentum parameter. Introducing this extra term can speed up convergence[87]. Figure 5.8 displays the effect of the momentum term qualitatively. All the parameters, α, β, and η should be chosen carefully in order to prevent networks from staying at a local minimum and to optimize the networks. We will investigate this matter later.

Recently, there has been a greater development of understanding of the interpretation of feed-forward neural networks[83, 89]. It appears that the output of a feed-forward neural network provides a direct approximation to the probability that a given feature vector X belongs to the signal class, under certain assump-

path through the mean
square error function
without momentum term

Space generated by the
mean square error
function

path through the mean square
error function with momentum term

Stable
Solution
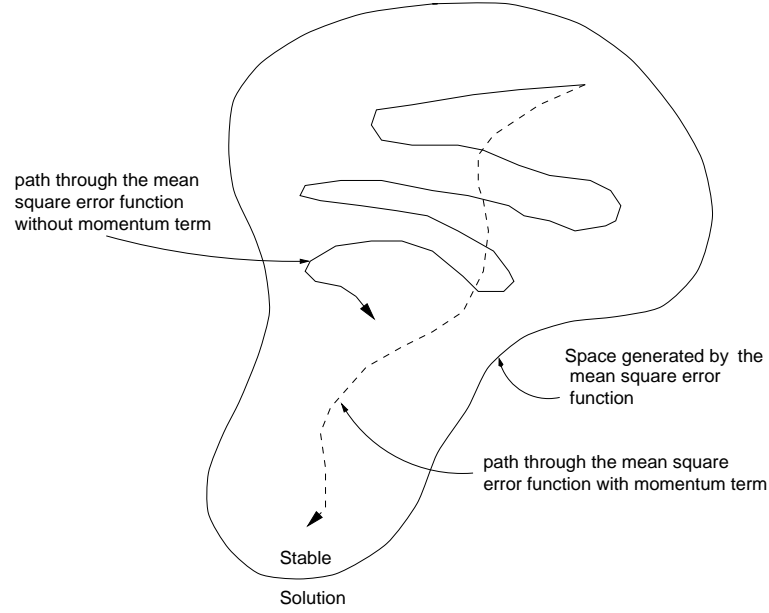
Figure 5.8: The addition of a momentum term during the training procedure.

tions. Because of the importance of this statement, we present the proof of this statement in the next section.

5.3.1 Probability and Neural Networks

Given a mixture of signal and background events, what is needed for the classification of an event is the optimal decision boundary in the parameter space. A possible way to construct such a decision boundary is to form a ratio of $P(S|X)$ to $P(B|X)$:

$$
\begin{aligned}
r &= \frac{P(S|X)}{P(B|X)} \\
&= \frac{P(X|S)P(S)}{P(X|B)P(B)},
\end{aligned}
\tag{5.13}
$$

where r is called the discriminant function, $P(S|X)$ is the probability that any event characterized by X is a signal event, $P(X|S)$ is the probability that a signal event has characteristics X (this is usually termed the likelihood function), and $P(S)$ is the *a priori* probability of an event being a signal event. In our case, $P(S)/P(B)$ is proportional to the ratio of the production and background cross sections, with the constraint that $P(S) + P(B) = 1$. A selection on r can, in principle, provide an optimal way to classify the event. It is often more convenient (and is analogous to the technique needed in covariance matrix analysis) to define the quantity:

$$
\begin{aligned}
f &= \log r \\
&= \log P(X|S) - \log P(X|B) + \log\left[P(S)/P(B)\right].
\end{aligned}
\tag{5.14}
$$

Now assuming that the likelihood functions can be represented as n-dimensional Gaussians, one can show that[90, 91]:

$$
\begin{aligned}
f &= \frac{1}{2}(X - \langle X \rangle)^T C_B^{-1}(X - \langle X \rangle) \\
&- \frac{1}{2}(X - \langle X \rangle)^T C_S^{-1}(X - \langle X \rangle) \\
&+ \frac{1}{2}\log\left(\frac{det C_B}{det C_S}\right) \\
&+ \log\left(\frac{P(S)}{P(B)}\right)
\end{aligned}
\tag{5.15}
$$

where $det(C)$ is the determinant of the matrix C. The function f is identical to Fisher's linear discriminant function introduced in the previous section, except for the last two terms. Therefore, it appears that Fisher's linear discriminant function is different from $\log r$ and does not contain the relative difference in a priori probabilities of the signal and the background.

Now, we turn to the connection between probability and an artificial neural network that has only a single output ($F \equiv F_i$). One can separate the mean square error function into two parts[90]:

$$E(F) = \frac{N_S}{N_p} \frac{1}{N_S} \sum^{N_S} (F - s)^2 + \frac{N_B}{N_p} \frac{1}{N_B} \sum^{N_B} (F - b)^2, \qquad (5.16)$$

where $N_p = N_S + N_B$, N_S is the number of signal events, N_B is the number of background events, and target values for signal and for background are s and b, respectively. If we consider the limit of $N_p \rightarrow \infty$, then the fractions N_S/N_p and N_B/N_p go over to $P(S)$ and $P(B)$, respectively, and sums go over integrals as:

$$E(F) = P(S) \int dX P(X|S)(F - s)^2 + P(B) \int dX P(X|B)(F - b)^2. \qquad (5.17)$$

where $P(X|S)$ and $P(X|B)$ appeared as a priori distributions of X for signal and background, respectively. Therefore, in the limit $N_p \rightarrow \infty$, one can regard the mean square error function E, as integrals of probabilities. Using the relations on conditional probabilities[91], namely:

$$
\begin{aligned}
P(S|X) &= P(X|S)P(S)/P(X) \\
P(B|X) &= P(X|B)P(B)/P(X) \\
P(X) &= P(X|S)P(S) + P(X|B)P(B), \qquad (5.18)
\end{aligned}
$$

we can rewrite the mean square error function as

$$
\begin{aligned}
E(F) &= \int dX P(X) \left[F^2 - 2FG(X, a, b) \right] \\
&+ \int dX P(X) \left[s^2 P(S|X) + b^2 P(B|X) \right], \qquad (5.19)
\end{aligned}
$$

where the function $G(X, a, b)$ is defined as

$$G(X, a, b) \equiv sP(S|X) + bP(B|X). \tag{5.20}$$

By factorizing the square term in $E(F)$ and rearranging the remaining terms we end up with the the following formula:

$$E(F) = \int dX P(X)(F - G)^2 + \int dX P(X)(s - G)(G - b). \tag{5.21}$$

Because our goal was to minimize $E(F)$ by adjusting ω and θ, and the second term has no dependence on F, the part relevant to the neural network must be the first term in the above equation. If a function F can be found to satisfy $F = G$, then the network output can be interpreted as:

$$F = sP(S|X) + bP(B|X). \tag{5.22}$$

The existence of such a function is always guaranteed because of a mathematical theorem[92]. The importance of a feed-forward neural network is that it is an example of such a function that minimizes $E(F)$ through a back-propagation technique[89]. Now, using the relation $P(S|X) + P(B|X) = 1$, one can derive the following relations:

$$
\begin{aligned}
P(S|X) &= \frac{(G - b)}{(s - b)}, \\
P(B|X) &= \frac{(s - G)}{(s - b)},
\end{aligned}
\tag{5.23}
$$
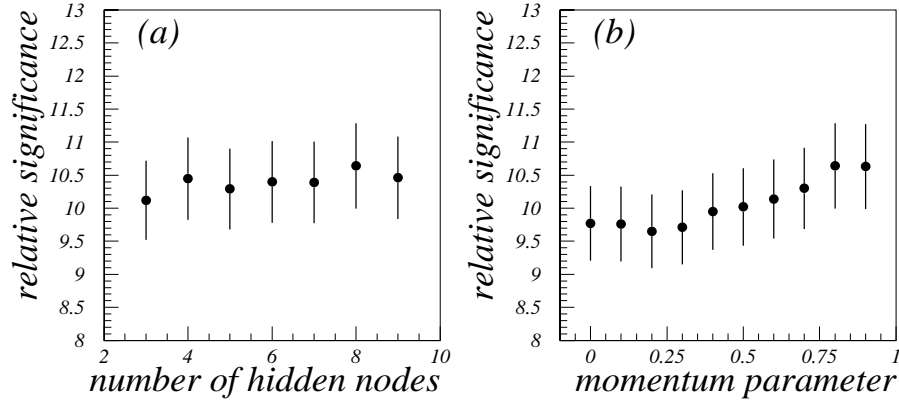
Figure 5.9: Relative significance of the signal from a mixture of 5000 signal and 5000 background events. The network outputs were required to be larger than 0.7. (a) Relative significance (signal/background) as a function of the number of hidden nodes, (b) relative significance as a function of the momentum parameter.

and find that the discriminant function r is given by:

$$r(G) = \frac{(G - b)}{(s - G)}. \qquad (5.24)$$

Therefore, classification based on the discriminant function can be approximated by using artificial neural networks. If one sets s to be unity and b to be zero, then

$$F = P(S|X), \qquad (5.25)$$

which means that the network output is the probability of obtaining a signal, for some specified feature vector (or event). In fact, in the following sections, we will set s to unity and b to be zero in order to have a direct interpretation of the network output as the probability, $P(S|X)$.
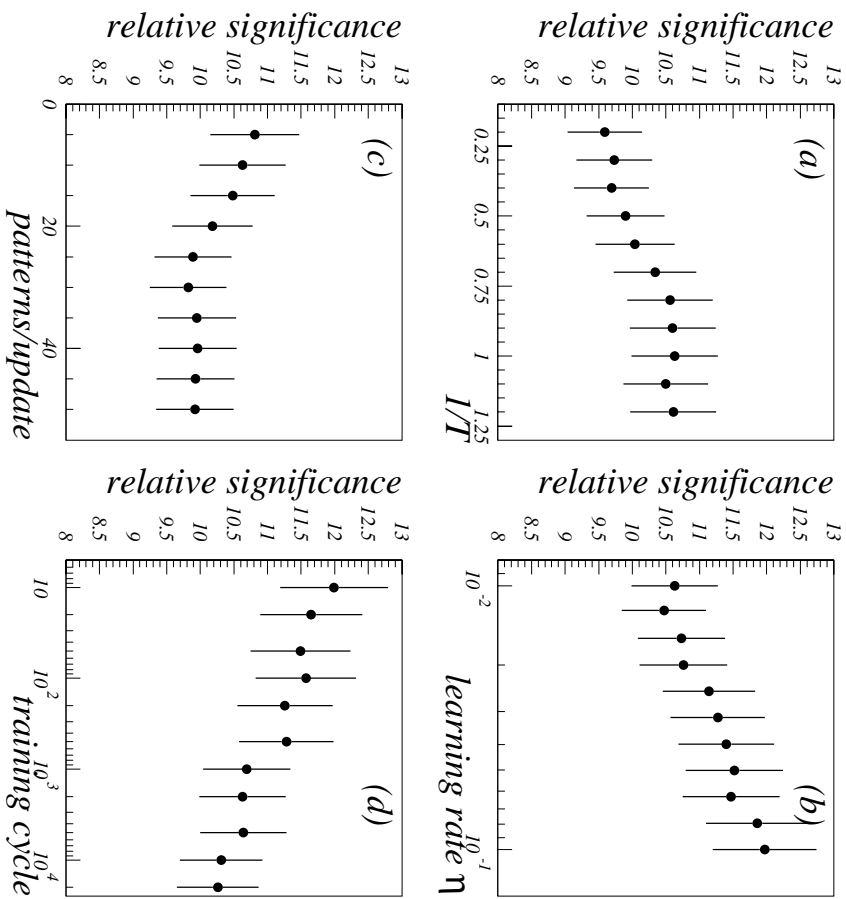
Figure 5.10: Relative significance of signal for network outputs larger than 0.7. (a) As a function of the inverse of the temperature (β), (b) as a function of the learning rate (η), (c) as a function of the number of patterns (events) prior to updating, (d) as a function of the number of training cycles.

5.3.2   Stability of Results from the Network

In order to make sure that the networks do not have any local minima in the network parameter space, we explored different regions of parameter space and different configurations of the networks (see below). We chose the feature parameters to be H_T^{3j}, $\mathcal{A}$, $\mathcal{C}$, $E_{T(5)}$, and H_T, and the following setting as defaults.

- network architecture: 5-8-1, namely, 5 input patterns, a layer of 8 hidden nodes, and a single output.

- momentum parameter α: set to 0.9 unless specified to the contrary.

- temperature (T): set to unity unless specified to the contrary.

- learning rate η: set to 0.01, unless specified to the contrary.

- number of patterns per update: this is the number of input patterns (feature vectors) to sum over prior to updating of the weights in the training of a given network, and is set to 10, unless specified to the contrary.

- number of training cycles: in the training procedure, many sets of event patterns are presented to the networks. This defines the number of training cycles, and is set to 2000 unless specified to the contrary.

The data sets used for training were identical to those used in training both the grid search and the covariance-matrix analysis. Networks were set to have one hidden layer, with s (target value for signal) defined to be unity, b (target value for background) to be zero, and to provide a single output. After the training, we tested 5000 signal events and 5000 background events, requiring the network output to be larger than 0.7, and investigated any changes in relative significance, defined by the number of surviving signal events divided by the number of surviving background events.

Before deciding on the parameters of the network, we varied the number of hidden nodes between three and nine. Figure 5.9 (a) shows the variations in the relative significance as a function of the number of hidden nodes defined in the network. The bars indicating statistical uncertainties. Within statistical uncertainty, the significance does not change with the number of nodes, indicating the stability of the network output against the number of hidden nodes. The greatest significance of the signal was obtained using the 5-8-1 configuration.

The momentum parameter α was varied from 0.0 to 0.9 in steps of 0.1. No indication of failures in training was observed, as shown in Fig. 5.9 (b). The largest significance was observed for $\alpha = 0.9$.

The inverse of the temperature β ($\beta=1/T$) was varied from 0.2 to 1.2 in steps of 0.1. Again, no indication of failures in training was observed, as shown in Fig. 5.10 (a). The largest significance was found for $\beta = 1.0$.

The effect of the learning-rate on the network performance was studied by setting its values from 0.01 to 0.1 in 10 uniform steps on a logarithmic scale. Figure 5.10 (b) shows the significance as a function of η. Although the largest significance was found for $\eta = 0.1$, a smaller value of η is preferred due to the fact that the setting did not allow the distribution of signal outputs to peak at 1.0 (assuming that we will cut the network output of the search sample close to unity later).

The effects of changing the number of event patterns per update, and of the total number of training cycle were investigated and are shown in Fig. 5.10 (c) and (d), respectively. Again, no indication of failure in training was detected. The training sample we used (finite number of patterns) may contain misleading regularities due to sampling. After a certain number of cycles the network can start to see these patterns to decrease the mean square error function[87]. This causes the overfitting of the network (known also as overtraining), and the decrease

of the significance in Fig. 5.10 (d) may be explained by this.

Considering the significance, as well as the distribution of the network output for the $t\bar{t}$ signal near 1.0, we chose the above default parameter values for our analysis.

5.3.3 Training

We used the default settings discussed in the previous section as values of parameters for the final training of the networks. It is our hope that the network method takes greater advantage of any correlations among parameters than for the case of our previous techniques.

Figure 5.11 (a) shows the distributions of network outputs for the signal (shaded area) and the background (black circles). The two distributions were normalized to the same area in order to compare their shapes. As we have argued, we can interpret the network output F as the probability of any given event being signal, $P(S|X)$. The signal populates the F values near unity, as opposed to background, which peaks near zero. As shown in Fig. 5.11 (b), if we consider the signal to background ratio prior to the muon-tagging, we are still at the level of $1/10$, even for the very highest network output. (The two distributions in Fig. 5.11 (b) are scaled by the number of events expected in the search sample, excluding 7219 training events.) When the two distributions are weighted by their respective tag-rates (for signal and for background), we see in Fig. 5.11 (c) that at the very highest network output we expect to have a signal to background ratio close to $\sim$ 1. This is comparable to the result obtained using covariance-matrix analysis (see Fig. 5.5). We will compare the performance of networks to other techniques later in this chapter. In Appendix A, we provide the result of the training process.
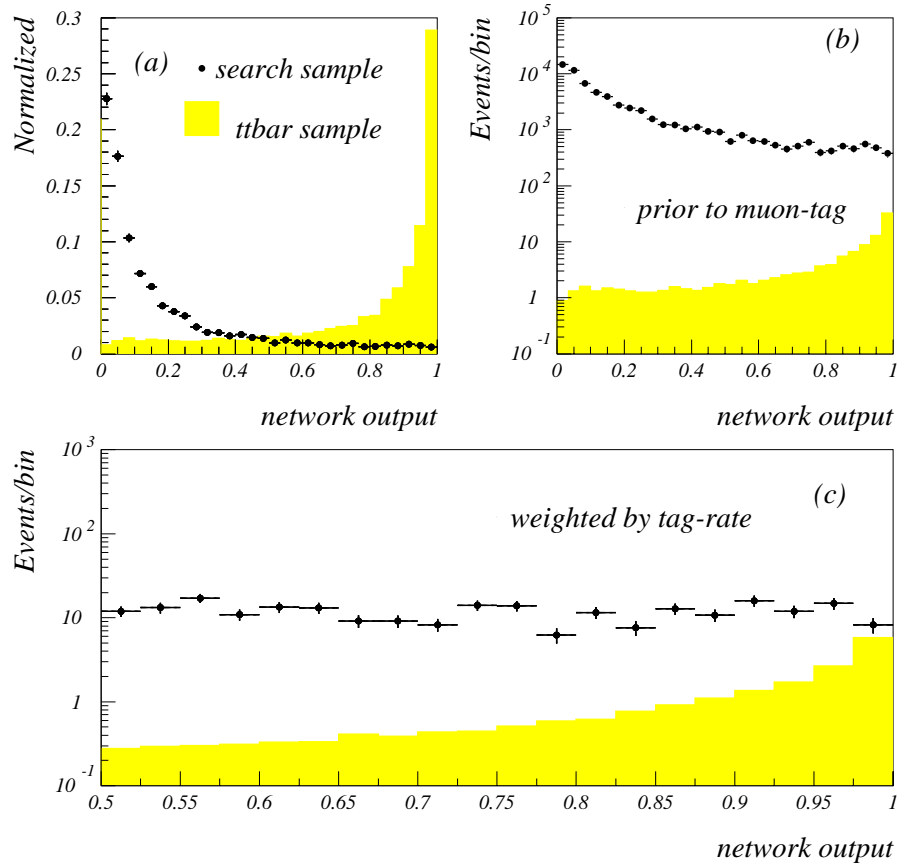
Figure 5.11: Results of the training of networks. (a) Distributions of network outputs using background (black circles) and signal (shaded area), normalized to the same area. (b) Distributions of network outputs, weighted by their respective production cross section. (c) Same as (b), but weighted further by the respective tag-rates.

Table 5.4: Dependence of results on the threshold used in the output of the neural network.

set	network output $>$	events in RUN I (tag not required)	expected N_{bkg} (tagged)	observed (tagged)	expected $N_{t\bar{t}}$ (tagged)
i	0.9	1248	44.4	58	10.9
ii	0.95	616	23.0	31	8.0
iii	0.96	480	18.2	24	7.2
iv	0.97	357	13.7	18	6.2
v	0.98	219	8.6	12	4.8
vi	0.99	90	3.8	4	2.8

5.3.4 Results

We used the network outputs of 0.9, 0.95, 0.96, 0.97, 0.98, and 0.99 as arbitrary choices of threshold values, and counted the number of background and signal events passing these thresholds. Again, we excluded the data used for training the networks in order to maintain the statistical independence of the final search sample. Table 5.4 lists the 6 threshold values used for network outputs, the number of untagged events in RUN I passing the cutoffs, the number of expected tagged background events, the number of observed tagged events, and the number of tagged $t\bar{t}$ events expected. Clearly, set **vi** is contained in set **v**, etc., and the observed numbers are not statistically independent. There is again a small excess of signal over the background, that will be discussed later in this chapter. More on the effect of selecting on network output will be given in Appendix A.

5.4 Comparison of Grid, Covariance Matrix, and Neural Network Analyses

Independent of the $t\bar{t}$ content of the search sample, we wish to compare the performance of the three different techniques. For the Grid method, the number of signal versus background events expected in the search sample was taken from Fig 5.2. For the covariance matrix analysis, we scanned the values of the Fisher's discriminant fuction from 2.2 to 8.6 in 30 steps, and counted the expected number of signal and background events. For the neural network analysis, we scanned values of the network output from 0.75 to 0.99 in 30 steps and also counted the expected number of signal and background events at each cutoffs. All results are shown in Fig. 5.12. The results from the covariance matrix analysis (dashed curve) are consistently below the optimal boundary defined by the grid search. We attribute this under-performance of the covariance matrix analysis relative to the grid search as due to: 1) the incompleteness of the symmetrization of the parameters and/or 2) non-linear correlations that are ignored and yield incorrect covariance matrices.

The results from the neural networks are given by the full curve in Fig. 5.12. In the region where one expects ~ 80 background events, the neural network provides the best discrimination of signal to background, and ~ 7 % better than the grid search.

5.5 Systematic Uncertainties

The following items comprise the major sources of systematic uncertainty that affect either the background estimate or the signal efficiency, and are common to all three techniques.
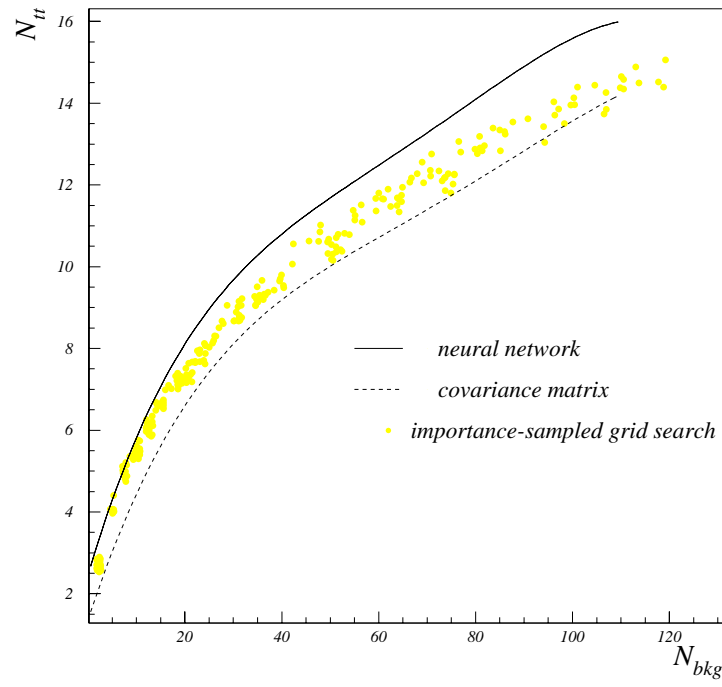
Figure 5.12: Comparison of signal/background response of the grid, covariance matrix, and neural network analyses.

- Because of the increase of multiple interactions with the instantaneous luminosity, the kinematic selection of events and the background estimation can depend on instantaneous luminosity. We estimated the uncertainty due to the increase of the multiple interactions by examining the muon tag-rate as a function of instantaneous luminosity and of the status of the multiple-interaction tool[93].* Figures 5.13 (a) and (b) show the tag-rate as a function of the instantaneous luminosity, and of the status of the multiple-interaction tool, for the Set 1 cutoffs in the grid search (Table 5.1). Although statistics are poor, we observe no clear indication of any change of tag-rate of "background" (i.e., data) with increasing luminosity (A linear fit to Fig. 5.13 (a) produced a slope of 0.00987 $\pm$ 0.00949.) The same is true, for example, for the Set **c** cutoffs in the covariance matrix analysis (see Table 5.3). This is shown in Figs. 5.13 (c) and (d). (A linear fit to Fig.5.13 (c) produced a slope of 0.00302 $\pm$ 0.0182.) Similarly, the output threshold of 0.9 in the neural-network analysis (shown in Figs. 5.13 (e) and (f)) shows no clear dependence on luminosity. (A linear fit to Fig. 5.13 (e) produced a slope of 0.000998 $\pm$ 0.00199.) We estimate ≤ 16 % uncertainty in the predicted background from dependence on luminosity for the grid, the covariance matrix, and the neural network analysis based on the one-sigma change of the result of the linear fit.

---

*Based on the probability of there being a single interaction, as determined through Level 0, calorimeter and the central detector (CD) information, this tool returns a value of -1 to 5. A value of 0 means that there was no interaction, -1 indicates that the CD did not find any vertices, but that other information points to a multiple interaction tool value of one; 1 means that there was "most likely" a single interaction; 2 means that it is "likely" that there was a single interaction; 3 means that it is "likely" that there was a multiple interaction; 4 means that it is "most likely" that there was a multiple interaction; and 5 means that it is "likely" that there were three or more interactions.
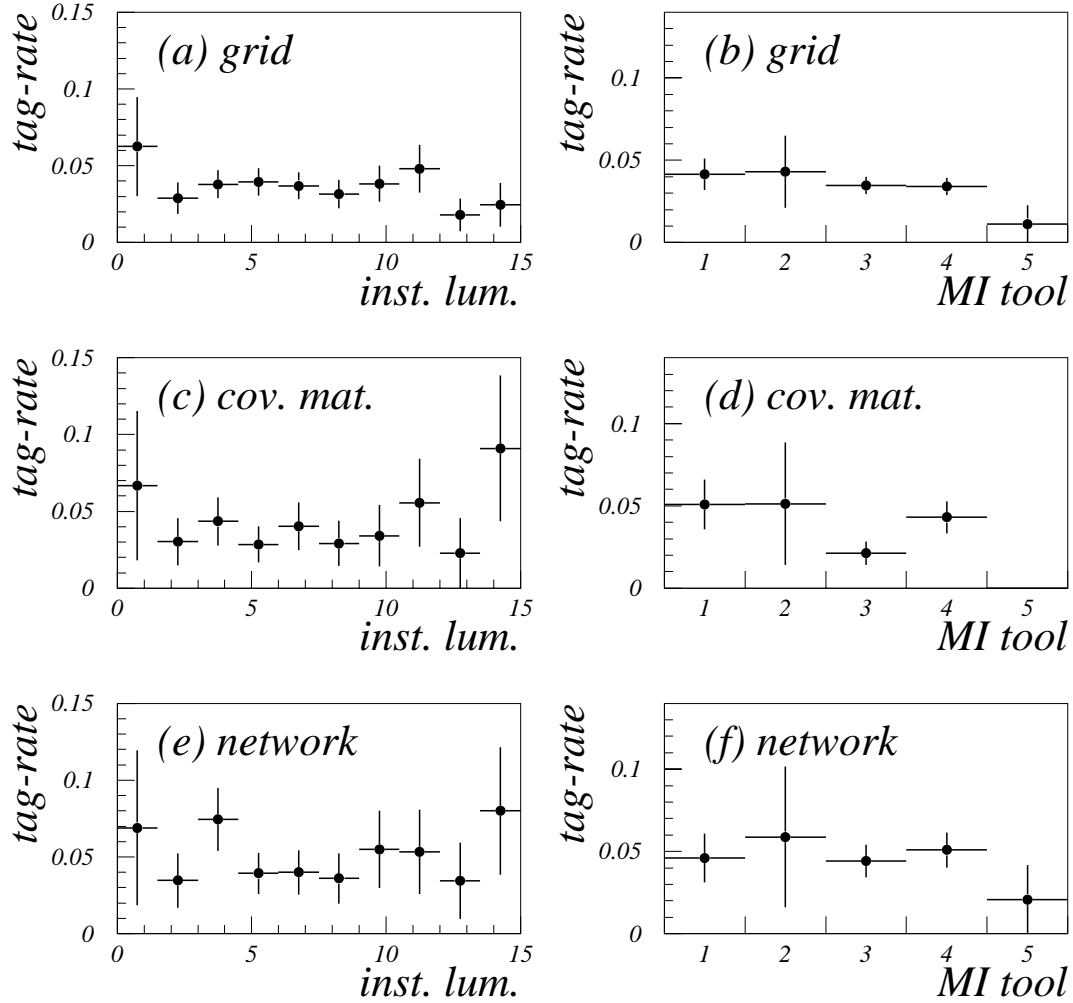
Figure 5.13: Studies of systematic uncertainties due to multiple interactions. (a) Muon tag-rate as a function of instantaneous luminosity (in units of 10^{30} events cm^{-2} s^{-1}) for the Grid method. (b) Muon tag-rate as a function of multiple interaction tool, for the Grid method, (c) same as (a) but for the covariance-matrix analysis, (d) same as (b) but for the covariance-matrix analysis, (e) same as (c) but for the neural network analysis, (f) same as (d) but for the neural network analysis.

- Leakage of hadrons from high momentum jets into the muon chambers may result in a false muon tag despite the thickness of the calorimeter (typically greater than 7 pion nuclear interaction lengths). High momentum hadrons or their hadronic spray into muon planes can be identified as muons, and we characterize them with unusually large energy deposits in the CH or ICD+MG sections of the calorimeter. Muons from π and K decays in the tracking region and in the calorimeter are included in the simulation of the background tagging rate and are not considered as leakage. From the dependence of the tag-rate on energy deposition in later parts of the calorimeter, we estimate a ≤ 2 % uncertainty in the predicted background due to the hadronic punch-through, assuming the leakage itself is not present in our data.

- The uncertainty in the parametrization of the tagging rate results in a 10 % uncertainty in the predicted number of background events. This was estimated by changing fit parameters in the tag-rate function (see Chapter 4) according to the one-sigma uncertainty of the result of fits.

- The b-quark content in QCD background can be affected by our kinematic selections. This can cause the background estimate to be inaccurate in regions where there is substantial signal. At present, we estimate a ≤ 10 % uncertainty in the predicted background from the effect of a variation of b-quark content with changes in our selection criteria[67].

- Events with a muon-tag are less likely to satisfy some fixed kinematic selection requirement because part of the energy of the tagged jet is taken by the muon and a neutrino. This can cause an error in the background estimate and in the signal efficiency. We estimate a 1 % uncertainty in the predicted

background and in the $t\bar{t}$ efficiency due to this possible bias[67].

- Any difference in the relative energy scale between data and Monte Carlo affects the efficiency for signal events. Varying the energy scale in the $t\bar{t}$ Monte Carlo by $\pm$ 5 % changes the efficiency for observing signal events by $\pm$ 10 % (for our standard selection criteria).

- The difference in the turn-on of trigger efficiency for JET_MULTI and for $t\bar{t}$ signal can affect the signal efficiency. The difference can originate from the modeling of electronic noise, or from the simulation of underlying events. We estimate $< 5\%$ uncertainty in signal efficiency from such sources[94].

- Uncertainty in parton fragmentation causes the major systematic uncertainty in signal efficiency. The uncertainty is estimated by comparing $t\bar{t}$ predictions from ISAJET and HERWIG generators. Figures 5.14 (a), (b), (c), and (d) show the fractional differences of efficiencies ($(\epsilon_{\mathrm{ISAJET}} - \epsilon_{\mathrm{HERWIG}})/\epsilon_{\mathrm{ISAJET}}$) for $\mathcal{A}$, H_T^{3j}, $E_{T(5)}$, and $\mathcal{C}$, respectively (again, for $m_t = 180$ GeV/c^2). The largest difference is observed in the distributions of H_T^{3j}. Based on that, we estimate 20% uncertainty in the overall signal efficiency.

- The constant tag-rate of 0.20 used for the signal assumes that the performance of all detector components was stable during the run. The changes in the muon tag-rate for background during the run ($\sim 20\%$) and the differences between $t\bar{t}$ Monte Carlo and data in the ICD and CH regions response have less then a 10 % effect[67] on the $t\bar{t}$ efficiency. We estimate a 20% uncertainty in the $t\bar{t}$ efficiency from any such changes in the muon-tag rate.

- The presence of $t\bar{t}$ events in the data sample used for estimating background must be taken into account. Calling the total number of events in the data
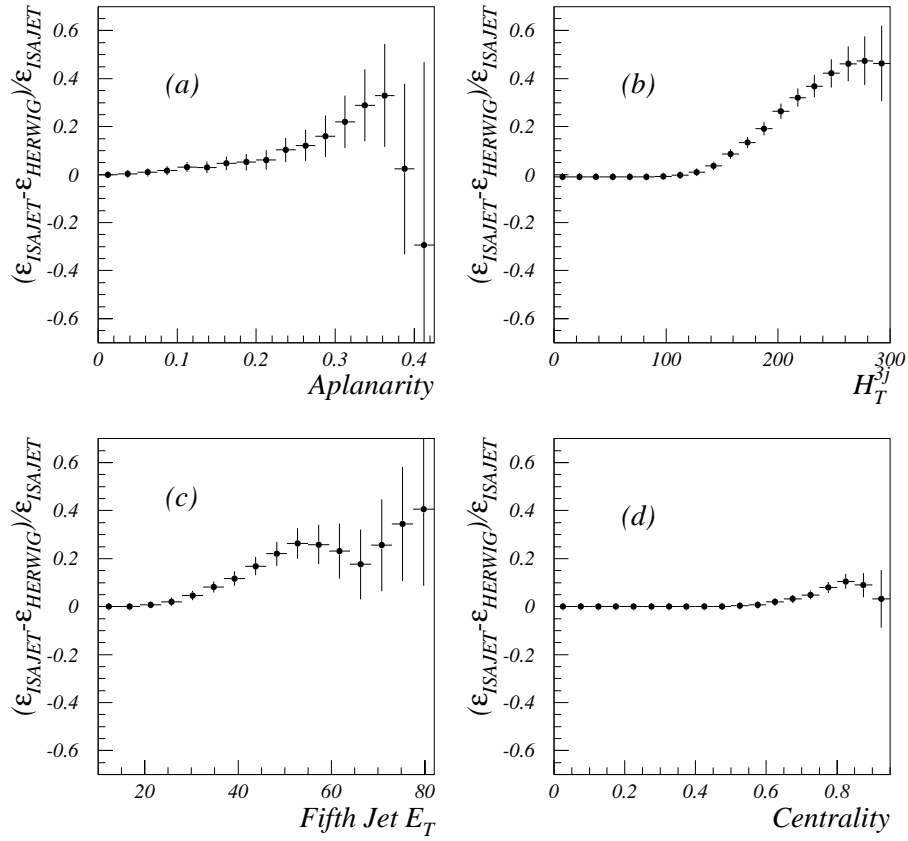
Figure 5.14: Fractional differences of efficiencies (ϵ) ((ϵ_{ISAJET} - ϵ_{HERWIG})/ϵ_{ISAJET}) for $m_t =$ 180 GeV/c². (a) Fractional differences in efficiencies as a function of threshold $\mathcal{A}$, (b) as a function of threshold H_T^{3j}, (c) as a function of threshold $E_{T(5)}$, (d) as a function of threshold $\mathcal{C}$.

Table 5.5: Summary of systematic uncertainties.

source	size of effect	affects
instantaneous luminosity	16 %	background estimation
leakage of hadrons to muon chambers	2 %	background estimation
fits to muon-tag rate	10 %	background estimation
b-quark content	10 %	background estimation
b-tag muons affect selection criteria	1 %	background estimation
	1 %	signal efficiency
jet energy scale	10 %	signal efficiency
trigger turn-on	5 %	signal efficiency
parton fragmentation	20 %	signal efficiency
constant tag-rate for signal	20 %	signal efficiency
$N_{bkg}(corr)$	10 %	background estimation

(given some set of selection criteria) prior to requiring a muon-tag N_{data}, we can define the fraction of signal events in the data (f_{sig}) sample as:

$$f_{sig} = \frac{N_{obs} - N_{bkg}}{0.2 N_{data}}$$

(5.26)

where the 0.2 corrects muon-tagged to untagged signal. The corrected background estimation therefore becomes:

$$N_{bkg}(corr) = N_{bkg}(1 - f_{sig})$$

(5.27)

This correction introduces a systematic uncertainty in the background estimate ~ 10 %[67].

The size of the above effects are summarized in Table 5.5. Among the sources that affect the background estimate, the uncertainty due to the multiple interactions is largest. The uncertainty in the assumption of a constant tag-rate for $t\bar{t}$ signal, independent of rapidity or E_T of the b-quark and uncertainty in parton

fragmentation have the greatest effect on $t\bar{t}$ efficiency.

5.6 Production Cross Sections

5.6.1 Significance and Standard Procedure for Cross Sections

We estimated the significance of any excess of $t\bar{t}$ signal relative to expected background for all sets of criteria developed in previous sections. We defined the probability (P) of seeing at least the number of observed events (N_{obs}), when only background is expected[32, 95]. The significance of any $t\bar{t}$ signal can be characterized by the likelihood of P being due to a fluctuation. If the distribution for the expected number of background events is assumed to be a Gaussian with mean b and the systematic uncertainty σ_b, then P can be calculated as:

$$
\begin{aligned}
P &= \sum_{n=N_{obs}}^{\infty} \int_0^{\infty} d\mu \frac{e^{-\mu}\mu^n}{n!} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(\mu-b)^2/2\sigma_b^2} \\
&= \int_0^{\infty} d\mu (1 - \sum_{n=0}^{N_{obs}-1}) \frac{e^{-\mu}\mu^n}{n!} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(\mu-b)^2/2\sigma_b^2}
\end{aligned}
\tag{5.28}
$$

The results of the calculation are listed in Tables 5.6, 5.7, and 5.8. Overall, the neural network provides greater significance than the grid or the covariance matrix technique, which provide similar results. Nevertheless, even for neural networks, the significance is too small to establish the existence of a $t\bar{t}$ signal in multi-jet final states.

Assuming that the excess of observed events is from $t\bar{t}$ production, we calcu-

Table 5.6: Cross sections for $t\bar{t}$ obtained from the importance-sampled Grid Search.

			Importance Sampled Grid Search		
set	N_{obs} (tagged)	$N_{bkg}(corr)$ (tagged)	Signal efficiency $\times$ BR (%)	P(background) Significance($t\bar{t}$)	Cross Section (pb)
1	97	89.2 ± 21.2	$\epsilon(160) = 2.72 \pm 0.83$ $\epsilon(180) = 3.64 \pm 1.11$ $\epsilon(200) = 4.36 \pm 1.33$	0.370 0.3 (s.d.)	3.8 ± 11.2 2.8 ± 8.4 2.3 ± 7.0
2	20	14.2 ± 3.4	$\epsilon(160) = 1.10 \pm 0.33$ $\epsilon(180) = 1.65 \pm 0.50$ $\epsilon(200) = 2.10 \pm 0.64$	0.148 1.0 (s.d.)	6.9 ± 7.0 4.6 ± 4.6 3.6 ± 3.7
3	8	6.9 ± 1.6	$\epsilon(160) = 0.76 \pm 0.23$ $\epsilon(180) = 1.14 \pm 0.35$ $\epsilon(200) = 1.61 \pm 0.49$	0.392 0.3 (s.d.)	1.9 ± 5.7 1.2 ± 3.8 0.9 ± 2.7

lated the cross section for the process using the following conventional formula:

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}(corr)}{\epsilon \times BR \times \mathcal{L}} \tag{5.29}$$

where $\epsilon \times BR$ is the branching ratio (BR) times the efficiency of the given a set of selection criteria for selecting $t\bar{t}$ events, and $\mathcal{L}$ is the total integrated luminosity. Because of the rejection of events with Main Ring activity and of the background-training sample (7219 events) from the search sample, the effective total integrated luminosity is 76.6 events/pb.

Table 5.6 lists the observed number of tagged events, the number expected from background, efficiency times branching ratios for three different input top mass values (denoted as $\epsilon(160), \epsilon(180)$ and $\epsilon(200)$, where numbers represent the input mass values in GeV/c^2), the values of P and the significances of any $t\bar{t}$ excess (in terms of standard deviations), and $t\bar{t}$ cross sections for three sets of selection criteria used in the grid-search technique. Uncertainties in the cross sections are

Table 5.7: Cross sections for $t\bar{t}$ obtained from the Covariance Matrix Analysis.

			Covariance Matrix Analysis		
set	N_{obs} (tagged)	$N_{bkg}(corr)$ (tagged)	Signal efficiency $\times$ BR (%)	P(background) Significance($t\bar{t}$)	Cross Section (pb)
a	149	131.9 $\pm$31.3	$\epsilon(160) = 3.08 \pm 0.94$ $\epsilon(180) = 4.05 \pm 1.23$ $\epsilon(200) = 4.90 \pm 1.49$	0.370 0.3 (s.d.)	7.2 $\pm$ 14.4 5.5 $\pm$ 11.0 4.5 $\pm$ 9.1
b	81	73.6 $\pm$ 17.4	$\epsilon(160) = 2.29 \pm 0.70$ $\epsilon(180) = 3.17 \pm 0.97$ $\epsilon(200) = 4.08 \pm 1.24$	0.354 0.4 (s.d.)	4.2 $\pm$ 11.2 3.1 $\pm$ 8.1 2.4 $\pm$ 6.3
c	43	41.0 $\pm$ 9.7	$\epsilon(160) = 1.63 \pm 0.50$ $\epsilon(180) = 2.38 \pm 0.72$ $\epsilon(200) = 3.25 \pm 0.99$	0.435 0.2 (s.d.)	1.6 $\pm$ 9.4 1.1 $\pm$ 6.4 0.8 $\pm$ 4.7
d	28	22.8 $\pm$ 5.4	$\epsilon(160) = 1.13 \pm 0.34$ $\epsilon(180) = 1.75 \pm 0.53$ $\epsilon(200) = 2.46 \pm 0.75$	0.249 0.7 (s.d.)	6.0 $\pm$ 9.0 3.9 $\pm$ 5.8 2.8 $\pm$ 4.1
e	15	13.0 $\pm$ 3.1	$\epsilon(160) = 0.79 \pm 0.24$ $\epsilon(180) = 1.23 \pm 0.38$ $\epsilon(200) = 1.86 \pm 0.57$	0.355 0.4 (s.d.)	3.4 $\pm$ 8.3 2.1 $\pm$ 5.3 1.4 $\pm$ 3.5
f	14	7.1 $\pm$ 1.7	$\epsilon(160) = 0.51 \pm 0.16$ $\epsilon(180) = 0.86 \pm 0.26$ $\epsilon(200) = 1.34 \pm 0.41$	0.0322 1.8 (s.d.)	17.6 $\pm$ 11.8 10.5 $\pm$ 7.0 6.7 $\pm$ 4.5
g	7	4.6 $\pm$ 1.1	$\epsilon(160) = 0.32 \pm 0.10$ $\epsilon(180) = 0.59 \pm 0.18$ $\epsilon(200) = 0.96 \pm 0.29$	0.203 0.8 (s.d.)	9.8 $\pm$ 12.2 5.3 $\pm$ 6.6 3.2 $\pm$ 4.0

estimated by adding statistical and systematic uncertainties in quadrature. The uncertainties in the cross sections are always greater than 90 % of the values of the cross section. The changes in efficiencies as a function of input top mass reflect the sensitivity of the selection criteria to the input mass value m_t.

Table 5.7 and 5.8 lists the analogous results that appeared in Table 5.6, but, respectively, for the covariance-matrix and the neural-network analyses. Uncertainties in the cross sections are treated in the same way as in the grid search.

As is demonstrated in the results of three independent analyses, we observe a

Table 5.8: Cross sections for $t\bar{t}$ obtained from the Neural Network Analysis.

	Neural Network Analysis				
set	N_{obs} (tagged)	$N_{bkg}(corr)$ (tagged)	Signal efficiency $\times$ BR (%)	P(background) Significance($t\bar{t}$)	Cross Section (pb)
i	58	42.0± 11.2	$\epsilon(160) = 2.24 \pm 0.78$ $\epsilon(180) = 3.03 \pm 1.06$ $\epsilon(200) = 3.76 \pm 1.32$	0.119 1.2 (s.d.)	9.4 ± 8.4 6.9 ± 6.2 5.6 ± 5.0
ii	31	21.5±5.7	$\epsilon(160) = 1.57 \pm 0.55$ $\epsilon(180) = 2.22 \pm 0.78$ $\epsilon(200) = 2.86 \pm 1.00$	0.116 1.2 (s.d.)	7.9 ± 7.1 5.5 ± 5.0 4.3 ± 3.9
iii	24	17.1±4.6	$\epsilon(160) = 1.38 \pm 0.48$ $\epsilon(180) = 1.99 \pm 0.70$ $\epsilon(200) = 2.58 \pm 0.90$	0.150 1.0 (s.d.)	6.5 ± 6.7 4.5 ± 4.6 3.5 ± 3.6
iv	18	13.0±3.5	$\epsilon(160) = 1.16 \pm 0.41$ $\epsilon(180) = 1.72 \pm 0.60$ $\epsilon(200) = 2.21 \pm 0.77$	0.180 0.9 (s.d.)	5.8 ± 6.4 3.9 ± 4.3 3.0 ± 3.4
v	12	7.9 ± 2.1	$\epsilon(160) = 0.87 \pm 0.30$ $\epsilon(180) = 1.32 \pm 0.46$ $\epsilon(200) = 1.76 \pm 0.62$	0.153 1.0 (s.d.)	6.1 ± 6.4 4.0 ± 4.2 3.0 ± 3.2
vi	4	3.8 ± 1.0	$\epsilon(160) = 0.45 \pm 0.16$ $\epsilon(180) = 0.77 \pm 0.27$ $\epsilon(200) = 1.10 \pm 0.38$	0.510 -	0.7 ± 6.5 0.4 ± 3.8 0.3 ± 2.7

small but consistent excess of $t\bar{t}$ signal in the multi-jet final states, and cross section values consistent with previous measurements in other modes of $t\bar{t}$ decay[18].

5.6.2   Cross Sections using Bayesian Theory

We also calculated the $t\bar{t}$ production cross section in an alternative way, basing it on Bayesian probability theory[96, 97]. The procedure is similar to that used in recent applications in other fields[98, 99, 100, 101] as well as in high energy physics[102, 103, 104, 105].

The total expected number of events (μ) is related to the production cross section ($\sigma_{t\bar{t}}$), the efficiency for observing the signal (including the branching ratio, ϵ), the integrated luminosity ($\mathcal{L}$), and the number of expected background events (b), as follows:

$$\mu = \mathcal{L}\epsilon\sigma_{t\bar{t}} + b. \tag{5.30}$$

The probability of observing n events, given an expectation value of μ, is assumed to be Poisson-distributed:

$$P(n|\mu, I) = \frac{e^{-\mu}\mu^n}{n!} \tag{5.31}$$

or

$$P(n|\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b, I) = \frac{e^{-(\mathcal{L}\epsilon\sigma_{t\bar{t}}+b)}(\mathcal{L}\epsilon\sigma_{t\bar{t}} + b)^n}{n!} \tag{5.32}$$

where I indicates all the information used to construct μ, as well as the assumption that n follows the Poisson distribution[106]. Note that, given a model for $t\bar{t}$ production, this probability can be used as a likelihood function for extracting parameters from the data. ($P(n|\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b, I)$ is often called the likelihood for the data.) Maximizing this likelihood is yet another way of the calculating cross section, that is, in fact, what was done in Reference [107]. However, one should recognize that the above likelihood is the probability of observing n events, given

μ, not a probability of μ being correct, given n. Therefore, what is, in fact, needed is the *inverse* of the likelihood, that is, $P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | n, I)$.

The inverse of the likelihood (or *a posteriori* probability) can be obtained using a basic theorem in probability theory, or Bayes' theorem[108] (also called the Principle of Inverse Probability), that relates the a posteriori probability to the likelihood and to the prior knowledge of the parameters (known as *a priori* probabilities). For our case, Bayes' theorem states that:

$$P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | n, I) \propto \frac{e^{-(\mathcal{L}\epsilon\sigma_{t\bar{t}}+b)}(\mathcal{L}\epsilon\sigma_{t\bar{t}} + b)^n}{n!} P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | I), \qquad (5.33)$$

where the constant of proportionality is determined by the fact that the right hand side of the above equation is a probability density function that must integrate to unity:

$$\int_0^\infty d\mathcal{L} \int_0^1 d\epsilon \int_0^\infty d\sigma_{t\bar{t}} \int_0^\infty db P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | n, I) = 1. \qquad (5.34)$$

Since we are not interested in probability distributions of $\mathcal{L}$, ϵ, or b, we can remove them. This is done naturally using a sum rule of probability theory[97]:

$$P(\sigma_{t\bar{t}} | n, I) = \int_0^\infty d\mathcal{L} \int_0^1 d\epsilon \int_0^\infty db P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | n, I). \qquad (5.35)$$

The a priori probability $P(\sigma_{t\bar{t}}, \mathcal{L}, \epsilon, b | I)$ represents any available information regardless of the outcome n (such as knowledge of the integrated luminosity within some estimated uncertainty, or the background within some uncertainty, etc.). For the calculation of the a priori probabilities in this analysis, we assume that all, except for $\sigma_{t\bar{t}}$ ($\mathcal{L}$, ϵ, and b), are uncorrelated, Gaussian distributed, and have uncertainties that can be obtained from studies of systematic uncertainties, as was done in the previous section. (It should be stressed that all the a priori probabilities must not reflect any information extracted from the data.)
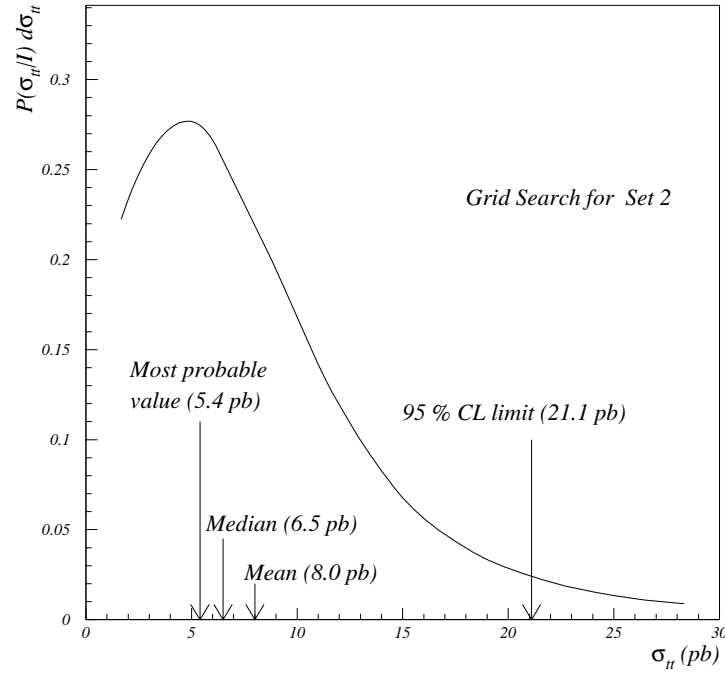
Figure 5.15: The a posteriori probability density function for $\sigma_{t\bar{t}}$, based on the result of cut set 2 of Grid Search.

Table 5.9: Cross sections for $t\bar{t}$ production using Bayesian theory, assuming $m_t = 180$ GeV/c^2.

	Importance Sampled Grid Search			
set	Most probable value (pb)	Mean value (pb)	Median value (pb)	95 % CL limit (pb)
1	$2.0 - 2.0 + 5.3$	9.8 ± 7.7	7.9	25.9
2	$5.4 - 4.3 + 5.2$	8.0 ± 6.5	6.7	21.1
3	$1.7 - 1.7 + 3.0$	5.8 ± 5.5	4.1	15.6
	Covariance Matrix Analysis			
a	$5.3 - 5.3 + 7.6$	12.6 ± 9.1	10.7	30.9
b	$3.0 - 3.0 + 5.5$	9.7 ± 7.8	7.8	25.7
c	$1.0 - 1.0 + 3.8$	7.5 ± 6.6	5.8	21.0
d	$3.8 - 3.8 + 4.8$	8.5 ± 6.8	6.9	22.5
e	$1.9 - 1.9 + 3.9$	7.4 ± 6.4	5.7	20.3
f	$10.3 - 8.4 + 7.5$	14.4 ± 8.3	12.9	31.0
g	$5.0 - 5.0 + 6.1$	10.7 ± 8.0	8.8	27.3
	Neural Network Analysis			
i	$6.3 - 6.3 + 5.9$	10.6 ± 7.5	9.0	25.8
ii	$5.4 - 5.4 + 5.0$	9.0 ± 6.8	7.5	22.6
iii	$4.7 - 4.5 + 4.6$	8.0 ± 6.4	6.5	20.7
iv	$4.1 - 4.1 + 4.3$	7.3 ± 6.0	5.9	19.1
v	$3.9 - 3.9 + 4.1$	7.6 ± 6.0	6.1	19.5
vi	$0.8 - 0.8 + 5.0$	10.8 ± 9.5	7.8	31.5

As an example, we consider the results of Set 2 from the Grid Search for $m_t = 180$ GeV/c^2. The calculated a posteriori probability as a function of the cross section $P(\sigma_{t\bar{t}}|n, I)$, is shown in Fig 5.15. The peak of the probability density function corresponds to the most probable cross section of 5.4 pb. The probability density function is asymmetric, and consequently the mean value of the distribution (8.0 pb) may not be a good estimator of $\sigma_{t\bar{t}}$. Integrating $P(\sigma_{t\bar{t}}|n, I)$ to the value of 0.95, we can set a 95 % upper confidence level (CL) limit for $\sigma_{t\bar{t}}$ of 21.1 pb. This procedure was repeated for all measurements assuming $m_t = 180$ GeV/c^2, and the results are summarized in Table 5.9. The a posteriori probability we used here is,

in fact, identical to the *extended* likelihood function[109] that one normally finds in the literature. Quoting the most probable value of the a posteriori probability distribution may therefore equivalent to applying the maximum likelihood technique. One difference, however, is that we integrated out parameters that are not of direct interest ($\mathcal{L}$, ϵ, and b) using the sum-rule of the probability theory, rather than extracting single values of $\mathcal{L}$, ϵ, and b that minimize the likelihood function.

Chapter 6

Extraction of the Mass of the Top Quark

The procedures used for extracting the mass of the top quark from the search sample can be divided into two parts. The first step involves the calculation of the mass value in each event, based on some kinematic fit or on any kinematic parameters that are correlated with the mass of the top quark. The second part involves performing a likelihood fit (or any other equivalent procedure) to all the data in order to extract the mass of the top quark. This chapter will describe these procedures.

6.1   Kinematic Fitting

As we discussed in Chapter 1, in multi-jet final states of $t\bar{t}$ production, one expects at least six jets in each event:

$$\overline{p} + p \rightarrow \quad t \qquad + \quad \overline{t} \qquad + \quad X$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$W^+ \quad + \quad b \qquad W^- \quad + \quad \overline{b}$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$2\,jets \qquad jet \qquad 2\,jets \qquad jet$$

where X represents the objects not directly involved in the $t\overline{t}$ production mechanism (commonly called *spectator* jets). There may also be additional jets from gluon radiation that we ignore at this stage. Once the jets in an event are identified, an invariant mass of three jets from the decay products of the top and antitop quarks can provide an estimate of the mass of the top quark. However, because of energy resolution and the fact that jets carry little information other than their energy, the challenge is to reduce background from wrong assignment of jets to the $t\overline{t}$ decay hypothesis. For the case of only six jets in an event, there are in total 90 different combinations that can, in principle, satisfy the $t\overline{t}$ production and decay hypothesis. In the presence of more than six jets in an event (from gluon radiation), the number of combinations is given by:

$$number\ of\ combinations\ for\ n\,jets = \frac{6!}{2 \times 2 \times 2} \times_n C_6 \qquad (6.1)$$

where n is the number of jets and $_n C_6$ the binomial coefficient. We see that, as n increases, the number of combinations increases rapidly (90, 630, 2520, ... for $n = 6,7,8$ ...), and the probability for correlating the correct jets with their parent sources must decrease.

6.1.1 Constrained Fitting

The kinematic fitting procedure used in our study consists of constructing a χ^2 based on the $t\bar{t}$ production and decay hypothesis, and minimizing the χ^2, subject to kinematic constraints (see below). The χ^2 can be written schematically as[110]

$$\chi^2 = (\mathbf{x} - \mathbf{m})^{\mathbf{T}}\mathbf{G}(\mathbf{x} - \mathbf{m}), \qquad (6.2)$$

where $\mathbf{x}$ and $\mathbf{m}$ denote vectors for the fitted variables and measured variables, respectively, and $\mathbf{G}$ denotes the inverse of the square of the error matrix. The fitting procedure allows measured variables to be pulled, subject to their uncertainties and to the kinematic constraints (introduced in Eq. (6.2) via LaGrange multipliers[110]) until the χ^2 reaches a minimum. The kinematic constraints, through the full error matrix, improve the resolution as a result of the fitting process. Because there are many successful fits for any given event, there is often no clear-cut way to select the best solution. We choose the combination of jets that gives the smallest fit χ^2. This frequently provides wrong assignment of jets, which broadens the experimental resolution of the mass of the top quark.

The constraints on energy-momentum conservation yield four equations:

$$E_{in} = \sum_{i=1}^{6} E_{(i)} + E_X \qquad (6.3)$$

$$\vec{p}_{in} = \sum_{i=1}^{6} \vec{p}_{(i)} + \vec{p}_X \qquad (6.4)$$

where the index i refers to jets in an event and X refers to the spectator jets. The invariant mass of two jets assigned as the decay products of W^+ (and W^-) boson is constrained to the known W-boson mass (m_W):

$$m_W^2 = \left(E_{(a)} + E_{(b)}\right)^2 - \left(\vec{p}_{(a)} + \vec{p}_{(b)}\right)^2 \tag{6.5}$$

where a and b refer to the jets from the decay of two W bosons. This provides two additional constraints. The three-jet invariant mass from t should equal the three-jet mass from $\bar{t}$, which is one more constraint,

$$\left(\sum_{c=1}^{3} E_{(c)}\right)^2 - \left(\sum_{c=1}^{3} \vec{p}_{(c)}\right)^2 = \left(\sum_{c=4}^{6} E_{(c)}\right)^2 - \left(\sum_{c=4}^{6} \vec{p}_{(c)}\right)^2, \tag{6.6}$$

where jets with indices $c=1,2,3$ are from the top quark and with indices $c=4,5,6$ are from the antitop quark. There are therefore a total of seven equations of constraint (C) with four unknown quantities (four-momentum of X), which overconstrains the fit (to 3C). One can choose to ignore constraints (Eq. 6.5) to the W boson mass, and replace it with the constraint that the two of two-jet invariant masses corresponding to the jets from the W be equal. This would reduce the fit to 2C. Also, if the transverse momenta P_X^x and P_X^y are assumed to be known (from the values of missing transverse energy in the calorimeter), the problem becomes a 5C fit.

SQUAW[110], a general-purpose kinematic fitting program that incorporates multi-vertex topology is used for the analysis. The key element in SQUAW is the linearization of the problem in the minimization of the χ^2[111]. We parametrized the four-momentum of a jet in SQUAW in terms of its azimuthal angle (ϕ), its polar angle θ (using $\tan\left(\pi/2 - \theta\right)$ as the variable in the fit), its energy (E), and its mass (m).

Unless stated otherwise, the uncertainties (standard deviations) assigned to our

reconstructed jets were as follows[59, 75, 112]:

$$\delta\phi = 0.035 \qquad (6.7)$$
$$\delta\tan(\pi/2 - \theta) = 0.040/\sin\theta$$
$$\delta E = 1.2\sqrt{E}$$
$$\delta m/m = 0.5.$$

The relatively large sampling term (120 %) in the jet-energy resolution is a result of jet reconstruction and faulty association of jets with their parent partons in the multi-jet environments[59, 75]. Also, regardless of parton flavor, jet masses in ISAJET (after the full detector simulation) are found to be distributed between 5 and 18 GeV/c², where the RMS width of the jet mass distribution is approximately half of its mean value. This is the reason for assigning $\delta m/m = 0.5$. We studied the case that the mass of the light quark $(u, c, d$ and $s)$ is set to zero and the mass of the b-quark is set to 5 GeV/c² and found marginal degradation in resolution of the fitted mass distribution.

In order to minimize wrong combinations of jets in the fits, we required that the two-jet invariant mass corresponding to the W-boson be within 40 GeV/c² of m_W, and the three-jet masses corresponding to the two top quarks be within 100 GeV/c² of each other, prior to the fit.

6.1.2 Parton-level Study

The performance of the kinematic fitting algorithm was first tested using events generated with ISAJET and HERWIG Monte Carlo programs at an input top quark mass of 200 GeV/c², for the case of just six quarks from $t\bar{t}$ decay (without allowing gluons to be radiated).
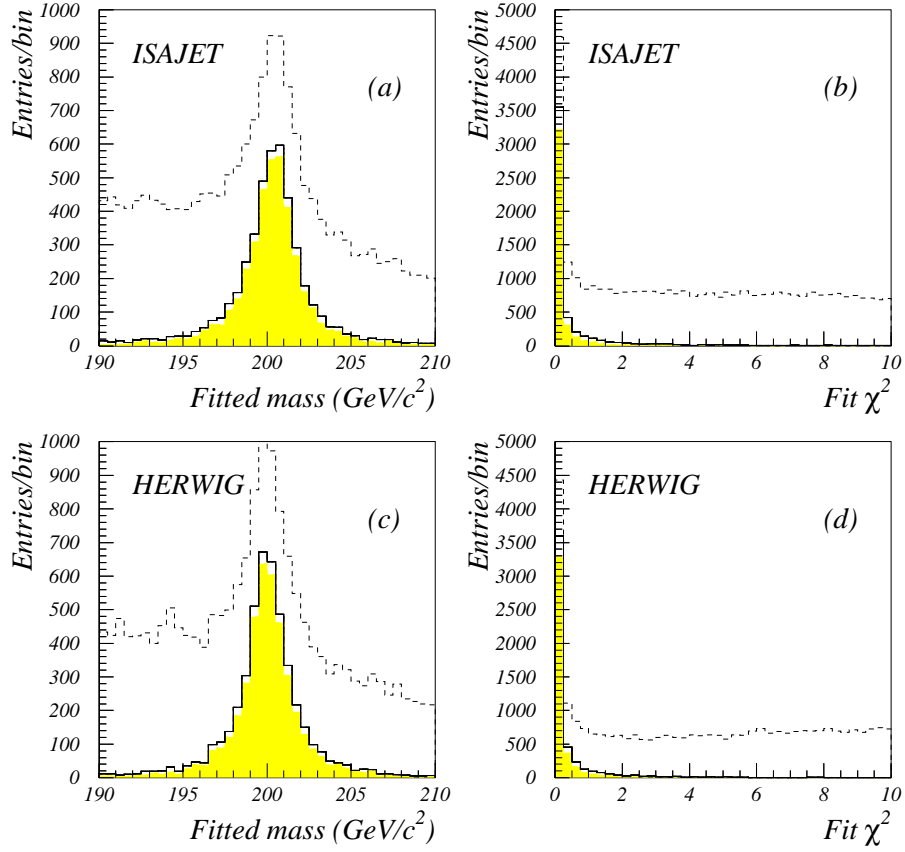
Figure 6.1: Mass and χ^2 distributions ($\chi^2 < 30$) for all fits (dashed), for solutions with the best-χ^2 (full lines), the subset of best-χ^2 solutions where jets are assigned correctly (shaded). (a) and (b) are calculated using ISAJET $t\bar{t}$ events, (c) and (d) are calculated using HERWIG $t\bar{t}$ events, both sets are for m_t=200 GeV/c^2, without presence of gluon radiation.

Subsequent to the fit, we required that the χ^2 for any fit to be less than 30 in order to be acceptable. We will refer "best-χ^2 solution" as the permutation of jets (or partons) that provides the minimum fit χ^2 in any event, and refer to the "correct combination" as the combination of jets (or partons) that corresponds to the right assignment of objects in the $t\bar{t}$ topology. Figure 6.1 shows the distributions in top-mass (m^{fit}) obtained in the fits, and the corresponding χ^2 distributions, for ISAJET and HERWIG $t\bar{t}$ events. The χ^2 distributions peak at zero, indicating that the $t\bar{t}$ hypotheses are indeed frequently satisfied. Full lines correspond to distributions when only best-χ^2 solutions are chosen, the shaded histograms are the ones when the best-χ^2 solution yields correctly assigned jets, and the dashed histograms are for all fits. The peaks in the fitted mass occur near the input mass value, 200 GeV/c^2, with an RMS of 3.9 $\pm$ 0.039 GeV/c^2 (full lines). HERWIG provides slightly a narrower (by $\sim$ 0.5 GeV/c^2) distribution in fitted mass (full lines).

We define fitting efficiencies ϵ as (all of numerators require $\chi^2 < 30$):

$$
\begin{aligned}
\epsilon(best\ \chi^2) &\equiv \frac{\texttt{\# of events with best} - \chi^2 \texttt{combination}}{\texttt{total \# of events}} \\[2mm]
\epsilon(corr.\ comb.) &\equiv \frac{\texttt{\# of events with correct combination}}{\texttt{total \# of events}} \\[2mm]
\epsilon(best\ and\ corr.) &\equiv \frac{\texttt{\# of events with best} - \chi^2 \texttt{ and correct comb.}}{\texttt{total \# of events}}
\end{aligned}
\tag{6.8}
$$

The values of the efficiencies are listed in Table 6.1, along with the mean values and RMS widths of the distributions in fitted mass. Efficiencies from HERWIG are somewhat higher than for ISAJET. At the parton level, both generators provide very promising results.

We define the pull quantities for the kinematic variables as[110]:

$$
\frac{(x_i - m_i)}{\sqrt{\langle (\delta x_i^2 - \delta m_i^2) \rangle}}
\tag{6.9}
$$
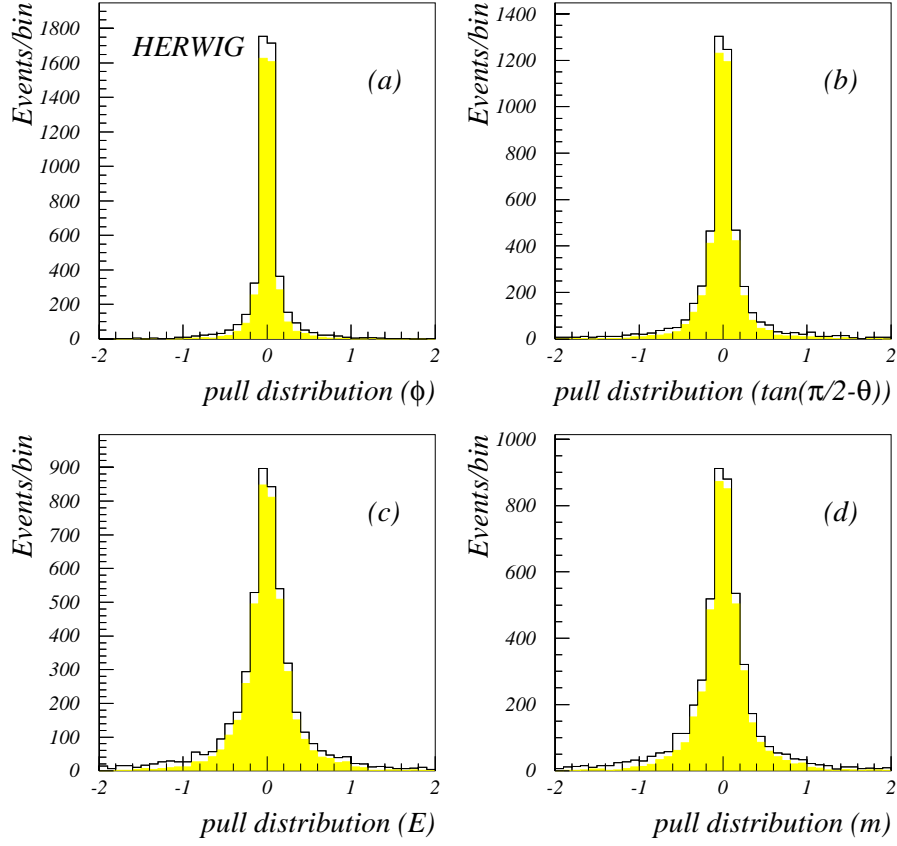
Figure 6.2: Distributions in pull quantities for HERWIG $t\bar{t}$ events with m_t=200 GeV/c^2. The histograms are for the best-χ^2 solutions, shaded are the subsets where the best-χ^2 solutions have correctly assigned jets. Pull distributions are for (a) ϕ, (b) $\tan(\pi/2 - \theta)$, (c) E, and (d) m.

Table 6.1: Results of $t\bar{t}$ fits at parton level, without gluon radiation.

		ISAJET	HERWIG
$\langle m^{fit} \rangle$ (GeV/c^2)	best χ^2	200.0 ± 0.056	199.8 ± 0.049
	corr. comb.	200.1 ± 0.058	200.0 ± 0.052
	best χ^2 and corr. comb.	200.2 ± 0.044	200.0 ± 0.038
Width (s.d.) (GeV/c^2)	best χ^2	3.9 ± 0.039	3.4 ± 0.035
	corr. comb.	4.0 ± 0.040	3.7 ± 0.037
	best χ^2 and corr. comb.	2.8 ± 0.031	2.5 ± 0.027
Efficiency (%)	ϵ(best χ^2)	99.7 ± 2.0	99.7 ± 2.0
	ϵ(corr. comb.)	97.4 ± 2.0	98.2 ± 2.0
	ϵ(best and corr.)	80.1 ± 1.7	83.2 ± 1.7

where m_i and δm_i are their values and uncertainties from the measurement, and x_i and δx_i are their values and uncertainties from the fit. If the input measurements and uncertainties are correctly assigned, and our hypothesis holds, then the pull quantities should have a mean of zero and a width of unity. Such pull quantities are therefore useful for checking the fitting procedure. Figures 6.2 (a), (b), (c), and (d) show pull distributions in ϕ, $\tan(\pi/2 - \theta)$, E, and m, respectively, using the HERWIG $t\bar{t}$ sample for $m_t = 200$ GeV/c^2. Here, again, best-χ^2 solutions are shown as the normal histograms, and best-χ^2 solutions that also have correct combinations are shown shaded. No biases are apparent in all the distributions. The widths of the distributions clearly do not equal unity, and are due to our using the incorrect (large) input uncertainties of Eq. (6.7) at the parton level.

6.1.3 Corrections to Jet Energy

As discussed in Chapter 3, the energy of a jet observed in the calorimeter is corrected using standard DØ procedures (CAFIX). This is based on balancing momentum in photon-jet and di-jet events, as well as on studies of minimum-bias jets for characterizing the underlying event. In addition, data from the test beam

are used to parametrize noise and showering characteristics of calorimeters[62]. However, CAFIX is not optimized for jet spectroscopy and reconstruction of masses from jets in a multi-jet environment. In particular, the reconstructed mass of a top quark is sensitive to gluon radiation, particle showerings (fragmentation), and showering in the calorimeter outside of a fixed cone of a jet. Consequently, an extra jet-energy correction (out of cone correction, or OOC) that is specific to the mass analysis of the top-quark was developed at DØ [113]. This correction is extracted from a linear fit of the parton energy versus the corresponding jet energy in HERWIG $t\bar{t}$ events with input top-quark masses ranging between 160 to 210 GeV/c^2, in steps of 10 GeV/c^2. The original studies were restricted to events that had one of the W-bosons decaying semi-leptonically, and jets satisfying $E_T >$ 15 GeV. The correction is parametrized as a function of E_T and η of jets that are already CAFIX-corrected (reconstructed using full detector simulation). We applied the same correction to our HERWIG $t\bar{t}$ all-jets sample of m_t=200 GeV/c^2 (all Ws decaying via quark pairs). Figure 6.3 shows the jet energy versus matched parton energy before (a) and after (b) applying this correction. The straight lines correspond to a 1:1 perfect energy scale. It is clear that the extra correction brings the jet energy closer to that of its initiating parton.

Figure 6.4 (a) and (b) show the reconstructed mass of the W boson and the mass of the top quark, respectively, for ISAJET $t\bar{t}$ events with m_t=200 GeV/c^2, using full detector simulation and jet reconstruction (but not kinematic fitting). Shaded, are distributions where jets are corrected just using CAFIX, and the normal histograms are distributions when both CAFIX and OOC corrections are applied. After the OOC corrections, the peaks of the distributions get closer to the input W mass (80.22 GeV/c^2) and the input top-quark mass (200 GeV/c^2), demonstrating the need for this extra correction after the CAFIX. Unfortunately, the OOC correction does not affect the fractional width of the distribution, as can be seen
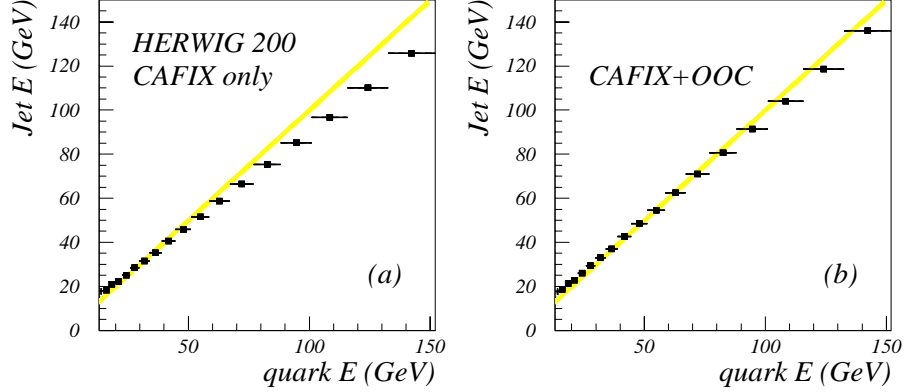
Figure 6.3: Comparisons of reconstructed jet energies (vertical axis) to those of their matched partons (horizontal axis), for HERWIG $t\bar{t}$ Monte Carlo, where (a) jets are CAFIX-corrected, (b) jets are CAFIX and OOC corrected.

from Table 6.2. This implies that the correction may not have great impact on mass resolution. Figures 6.4 (c) and (d) show the analogous distributions, but for HERWIG $t\bar{t}$ events. Table 6.2 indicates that the peaks of the mass distributions get somewhat closer to their input values, after the OOC corrections, but the fractional widths again remain unchanged. Similar conclusions were reached for checks performed at other input masses.

6.1.4 Jet-level Study

We repeated the test of the $t\bar{t}$ fitting algorithm using jet reconstruction in the fully simulated DØ detector environment, applying both CAFIX and OOC packages.

Figures 6.5 (a) and (b) show the distributions in top-mass and the corresponding χ^2 distributions, respectively, for fitted ISAJET $t\bar{t}$ events with m_t=200 GeV/c^2. The regular histograms are the distributions for best-χ^2 solutions, and the shaded histograms are the subset of best-χ^2 solutions where jets are correctly assigned.
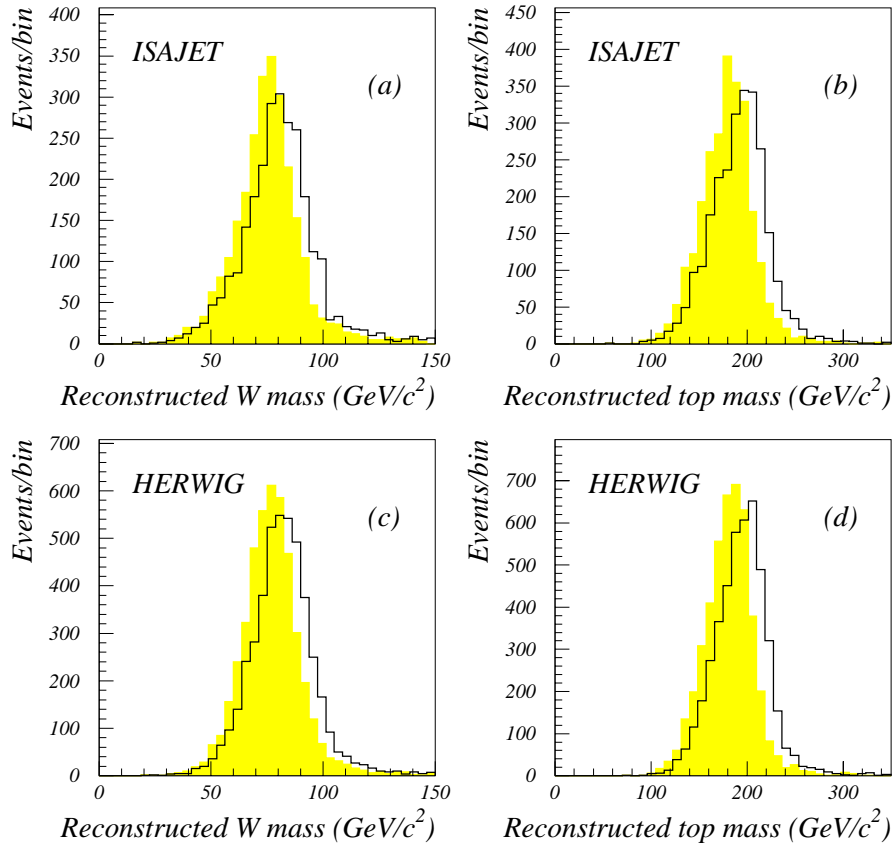
Figure 6.4: Reconstructed W and top-quark mass distributions. Shaded histograms correspond to distributions for jets that have had only CAFIX corrections, and the normal histograms to jets with both CAFIX and OOC corrections. (a) and (b) are for ISAJET $t\bar{t}$ events, and (c) and (d) are for HERWIG $t\bar{t}$ events, both samples for m_t=200 GeV/c^2.
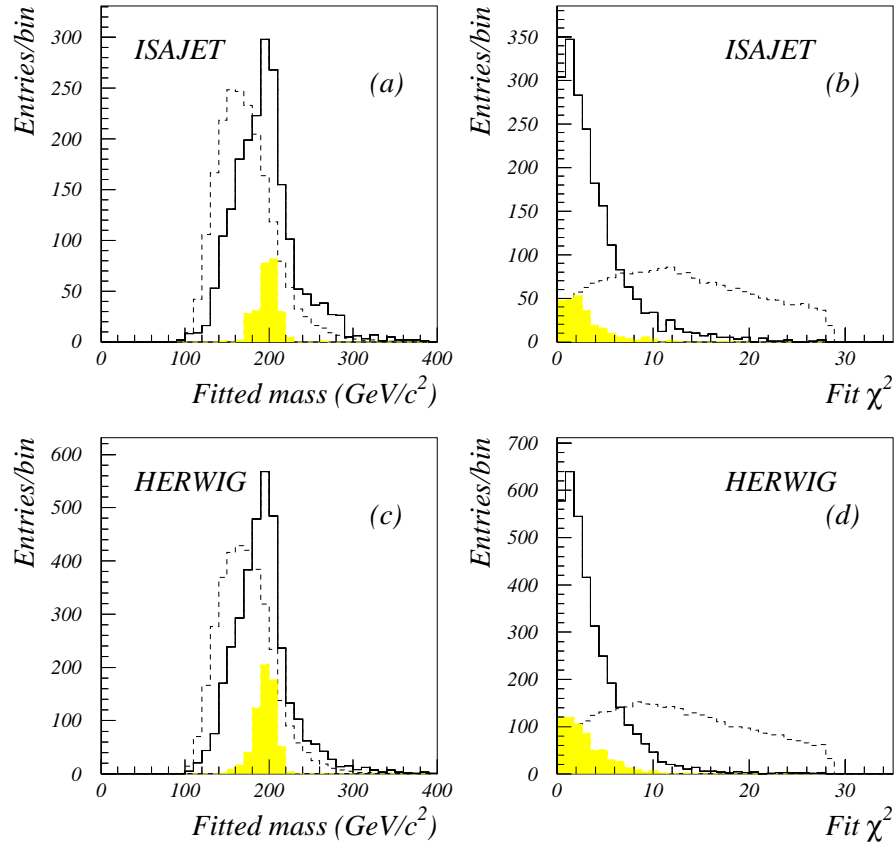
Figure 6.5: Mass and χ^2 distributions for all fits with $\chi^2 < 30$ (dashed, normalized to have same areas as normal histograms), for solutions with the best-χ^2 (full lines), the subset of best-χ^2 solutions where jets are assigned correctly (shaded). (a) and (b) are calculated using ISAJET $t\bar{t}$ events, (c) and (d) are calculated using HERWIG $t\bar{t}$ events, both sets are for m_t=200 GeV/c^2.

Table 6.2: Means and widths of the reconstructed jet masses of the W and of the top quark for ISAJET and HERWIG $t\bar{t}$ events.

		ISAJET	HERWIG
$(m_w^{true} - \langle m_w \rangle)/$	CAFIX only	6.0 ± 0.38	4.1 ± 0.25
m_w^{true} (%)	CAFIX+OOC	0.0 ± 0.41	-2.0 ± 0.27
$\sigma_{m_w}/\langle m_w \rangle$	CAFIX only	20.7 ± 0.30	17.7 ± 0.19
(%)	CAFIX+OOC	20.7 ± 0.30	17.7 ± 0.19
$(m_{top}^{true} - \langle m_{top} \rangle)/$	CAFIX only	10.5 ± 0.28	9.5 ± 0.19
m_{top}^{true} (%)	CAFIX+OOC	4.2 ± 0.20	3.1 ± 0.21
$\sigma_{m_{top}}/\langle m_{top} \rangle$	CAFIX only	15.7 ± 0.22	14.4 ± 0.15
(%)	CAFIX+OOC	15.7 ± 0.22	14.5 ± 0.16

The dashed histograms correspond to all solutions with $\chi^2 < 30$, normalized to have same area as the regular histograms. For the best-χ^2 combinations, the mean value of the fitted mass is lower than the input mass, by 5.4 ± 0.84 GeV/c^2, and has an RMS width of 37.7 ± 0.59 GeV/c^2. When jets are assigned correctly, the mean value is also smaller by 9.0 ± 0.61 GeV/c^2, and has an RMS width of 17.6 ± 0.43 GeV/c^2. It appears, therefore, that a major contribution to the large width is the low efficiency of obtaining the correct combinations for the solutions with best-χ^2. The fraction of correct combinations in solutions having best χ^2 is only 12.9 ± 0.94 % for ISAJET, as indicated by the areas of the respective histograms. Nevertheless, the χ^2 distribution peaks near zero, indicating that the uncertainties are not unreasonable and that the fitted hypotheses are well satisfied. Figures 6.5 (c) and (d) show the analogous distributions for HERWIG $t\bar{t}$ events with m_t=200 GeV/c^2. The HERWIG results are somewhat narrower in the fitted mass, with the solutions with with best-χ^2 combination having a mean of 192.0 ± 0.59 GeV/c^2 and an RMS width of 34.9 ± 0.42 GeV/c^2. The distribution for the correct combination has a mean of 190.3 ± 0.35 GeV/c^2 and an RMS width of 15.8 ± 0.25 GeV/c^2.

Events/bin

300

250

200

150

100

50

0

ISAJET

(a)

0   100   200   300   400

Top quark mass (GeV/c²)

Events/bin

600

500

400

300

200

100

0

HERWIG

(b)

0   100   200   300   400

Top quark mass (GeV/c²)
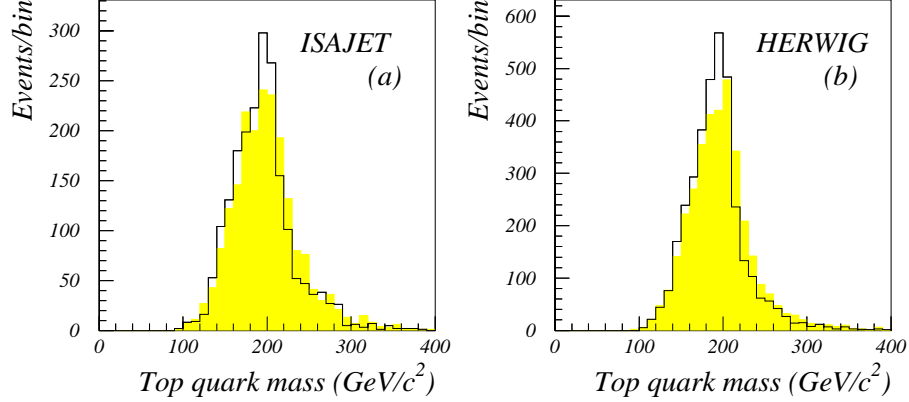
Figure 6.6: Mass distributions before (shaded) and after (normal) kinematic fitting, for ISAJET (a) and HERWIG (b) samples, both for $m_t = 200$ GeV/c².

The mean values and widths of the distributions of the fitted masses are listed in Table 6.3. By comparing the values of $\sigma_{m_{top}}/\langle m_{top}\rangle$ for the case of the correct combination in Table 6.3 and in Table 6.2, we see that kinematic fitting, in fact, reduces the fractional width of the mass distribution by ~ 40 %, in both ISAJET and HERWIG samples.

We also compared the 3-jet invariant mass distribution before and after the fitting, using for both cases the jet assignments chosen by the best-χ^2 solution. Figures 6.6 (a) and (b) show the 3-jet invariant mass distribution before (shaded histograms) and after the fit (normal histograms). In the case of the ISAJET sample, the RMS width of the distributions changes from 40.7 ± 0.91 (before the fit) to 37.7 ± 0.84 GeV/c² (after the fit). For HERWIG, the width changes from 38.0 ± 0.64 to 34.9 ± 0.59 GeV/c², as indicated in Fig. 6.6 (b). Because the statistical uncertainties in the widths are ≤ 1 GeV/c², these appear to be significant changes, and suggest that fitting is beneficial for improving mass resolution.

Figures 6.7 (a), (b), (c), and (d) show pull distributions for ϕ, $\tan(\pi/2 - \theta)$,

Table 6.3: Results of fits with fully reconstructed jets.

		ISAJET	HERWIG
$\langle m^{fit} \rangle$ (GeV/c^2)	best χ^2	194.6 ± 0.84	192.0 ± 0.59
	corr. comb.	191.0 ± 0.61	190.3 ± 0.35
	best χ^2 and corr. comb.	197.6 ± 0.84	195.0 ± 0.53
Width (s.d.) of m^{fit} (GeV/c^2)	best χ^2	37.7 ± 0.59	34.9 ± 0.42
	corr. comb.	17.6 ± 0.43	15.8 ± 0.25
	best χ^2 and corr. comb.	13.4 ± 1.9	13.3 ± 1.9
$(m_{top}^{true} - \langle m^{fit} \rangle)/$ m_{top}^{true} (%)	best χ^2	2.7 ± 0.42	4.0 ± 0.29
	corr. comb.	4.5 ± 0.31	4.9 ± 0.22
	best χ^2 and corr. comb.	1.2 ± 0.42	2.5 ± 0.26
$\sigma_{m_{top}}/\langle m^{fit} \rangle$ (%)	best χ^2	19.4 ± 0.32	18.2 ± 0.22
	corr. comb.	9.2 ± 0.23	8.3 ± 0.13
	best χ^2 and corr. comb.	6.8 ± 0.98	6.8 ± 0.99
Efficiency (%)	$\epsilon(\text{best } \chi^2)$	98.7 ± 3.1	99.0 ± 2.4
	$\epsilon(\text{corr. comb.})$	40.6 ± 1.7	57.4 ± 1.6
	$\epsilon(\text{best and corr. comb.})$	12.7 ± 0.84	17.9 ± 0.77

E, and m, respectively, using HERWIG $t\bar{t}$ events with $m_t = 200$ GeV/c^2. Again, best-χ^2 solutions are shown in the normal histograms, and best-χ^2 solutions with correct combinations are shown shaded. The curves show the result of Gaussian fits to the pulls. Although qualitatively the distributions are acceptable, the pulls in energy and mass of jets have means of -0.7 indicating a bias in the energy and mass scale of jets. The jet energies appear to be shifted to too low values as a result of the fitting process. This is not unexpected, considering the large uncertainties on the variables, and the assumption on linearity in the minimization procedure. Because the data and Monte Carlo are treated in same fashion, any shifts in mass can be recovered through the use of correct templates that map input to output mass. We attempted to correct the jet energy scale based on the pull distributions, but found no improvement in the resolution. Therefore, we decided to use the same correction adapted for lepton+jets analysis in DØ[113].
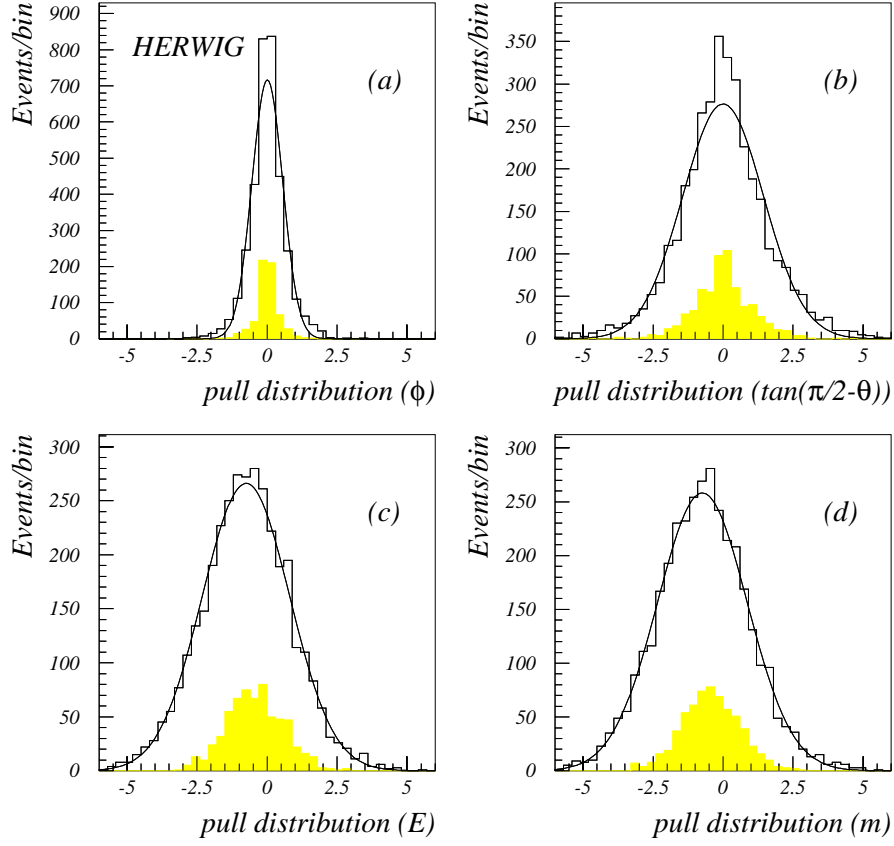
Figure 6.7: Distributions of pull quantities for HERWIG $t\bar{t}$ events with m_t=200 GeV/c^2. Histograms are for the solutions with best-χ^2, shaded areas refer to the subset of best-χ^2 solutions where jets are correctly assigned. The pull distributions are: (a) ϕ, (b) $\tan{(\pi/2 - \theta)}$, (c) E, and (d) m.
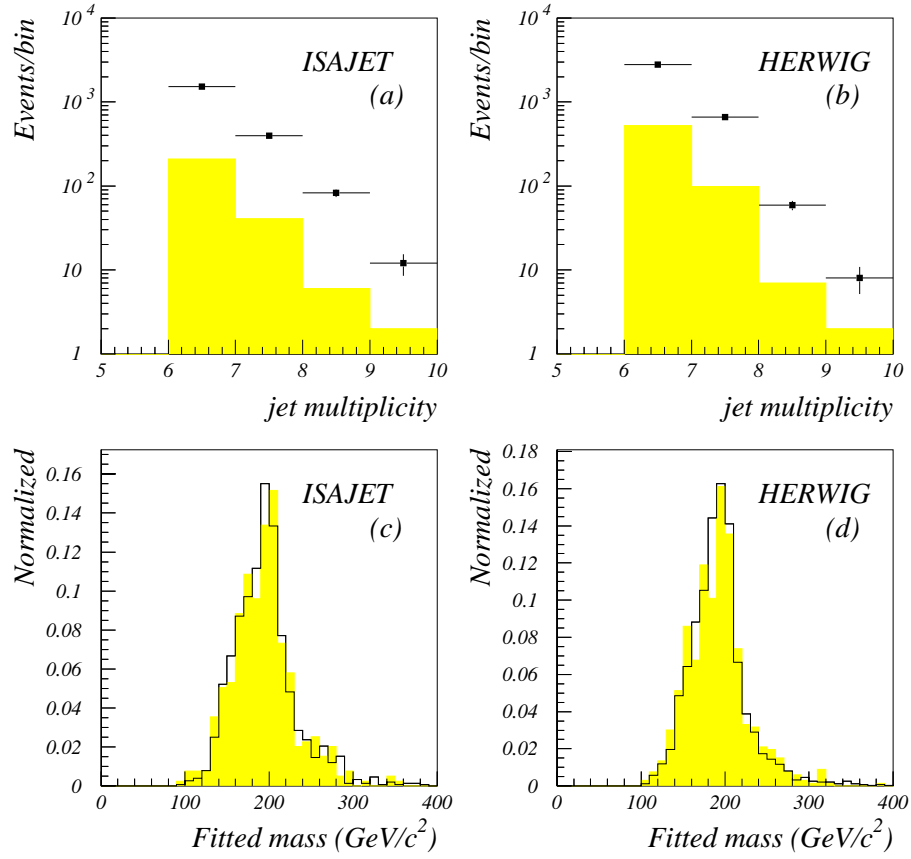
Figure 6.8: Jet multiplicity and mass distributions. Black squares in (a) and (b) are distributions for the parent samples (best-χ^2). Shaded histograms in (a) and (b) are the distributions when the solution with best -χ^2 also yields the correct combinations of jets. The normal histograms in (c) and (d) are distributions with events that have only 6 jets, and the shaded histograms are for 7 jets. Results in (a) and (c) are for ISAJET, and (b) and (d) are for HERWIG.

6.1.5 Effects of Extra Jets

For the case when there are more than 6 jets in an event, as we discussed previously, there are at least 630 possible combinations for the fit. Other studies have shown[32] that including these extra jets has very little impact on the mass resolution of the top quark. We therefore chose to use only the first six jets in our fits (where the jets are ordered in descending order in E_T). Figures 6.8 (a) and (b) show the jet multiplicity distributions for ISAJET and HERWIG $t\bar{t}$ events with $m_t = 200$ GeV/c^2. The black square points correspond to the multiplicities for the samples with best-χ^2, and the shaded histograms show the jet multiplicity for the solutions with best-χ^2 that have correct combinations of jets. The efficiency of finding the correct combination does not depend strongly on the jet multiplicity, indicating that our decision to use only the top 6 jets may have merit. Figures 6.8 (c) and (d) show the fitted mass distributions for a total of six (normal histograms) and seven jets (shaded), for ISAJET and HERWIG samples, respectively. The distributions have been normalized to the same area. For both generators, the RMS widths of the distributions get only little wider for the case of 7 jets (0.4 ± 1.5 GeV/c^2 for ISAJET and 1.3 ± 1.1 GeV/c^2 for HERWIG).

In order to explore the possibility of obtaining better performance from the fitting algorithm from merging extra jets in an event, we tested two such merging algorithms. First, the pair of jets with the smallest relative transverse momentum k_T (calculated from the projection of the two jet momentum vectors) were added together ("k_T-merging"). Figures 6.9 (a) and (b) show the fitted mass distributions using this k_T-merging (shaded histograms), together with the standard procedure (normal histograms) for ISAJET and HERWIG events, respectively. Unfortunately, k_T merging widens the RMS width of the distribution by 8 ± 2.1 % and 13 ± 1.9 % for ISAJET and HERWIG, respectively. The second algorithm involved merging
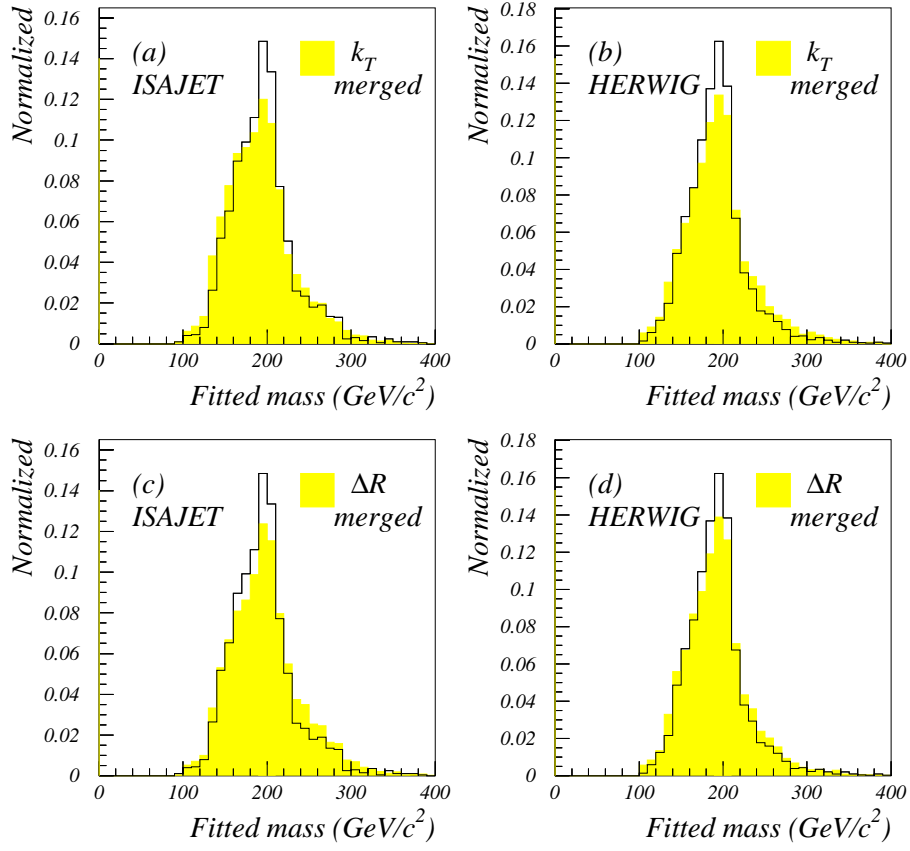
Figure 6.9: Effect of jet-merging on the distributions of fitted mass. The normal histograms show distributions using our standard procedure (just the 6 leading jets). Shaded histograms in (a) and (b) show the distributions applying k_T merging, and (c) and (d) applying with $\Delta \mathcal{R}$ merging.

of two jets when they were closer than a certain distance ($\Delta\mathcal{R} = 1.6$ was chosen for our studies) in η-ϕ space. The adding of four-momenta of such jets ("$\Delta\mathcal{R}$-merging"), also produced 9.7 ± 2.2 % and 8.4 ± 1.8 % degradation in the fitted mass resolution for ISAJET and HERWIG, respectively. This is shown in Figs. 6.9 (c) and (d). Since the merging of jets did not improve the mass resolution, in what follows we consider only the six leading jets and no merging.

Plotting the relation between the mean fitted mass and the input mass of the top quark, prior to applying any additional selection criteria beyond the χ^2 and the standard offline selections (selection criteria amount to search sample, see Chapter 4), one finds that the relationship is linear, with a slope 0.72 ± 0.014 and an offset of 49 ± 2.7 GeV/c^2 . This is shown in Fig. 6.10, where the black points were obtained for event samples with $m_t = 120$, 140, 160, 180, 200, and 220 GeV/c^2, using the ISAJET event generator. The dashed line is a result of a linear fit to the black circles, and the dotted line drawn along the diagonal (reflecting perfect response) is for comparison. The slope represents the effective sensitivity, or degradation in the extraction of the mass of the top quark in this analysis. The reasons that the slope does not equal unity are believed to be partially due to the H_T requirement in the offline selection, partially to the uncertainty in the energy scale, but mainly to the fitting algorithm itself, which often yields the wrong combination of jets for the best-fit χ^2. It is, of course, important to have selection criteria that do not degrade the slope, and also provide a good signal to background ratio. This will yield a smaller uncertainty on the extracted mass of the top quark from a likelihood fit that will be discussed in the following section.
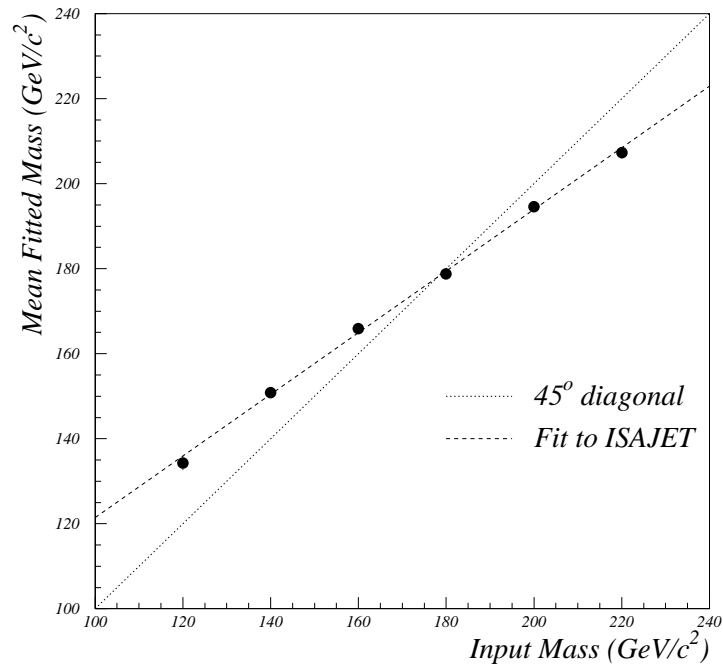
Figure 6.10: Mean fitted mass as a function of input mass value for samples of ISAJET $t\bar{t}$ events. The dashed line is the result of a linear fit to points, and the dotted line along the diagonal is for the comparison. (The statistical uncertainties are smaller than the size of the points.)

6.2   Approach using Likelihood Method

Once we specify a distribution that is sensitive to the mass of the top quark, and we have a model for the signal as a function of input mass, and a background available (in our case from data), we can perform a likelihood fit to extract the mass of the top quark. The procedure is described in references [32, 107, 114, 115], and uses the following "unbinned" likelihood function that is maximized with respect to its three parameters:

$$
\mathcal{L}(m_t, n_s, n_b) = \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(n_b - N_b)^2 / 2\sigma^2} \right] \left[\frac{(n_s + n_b)^N e^{-(n_s + n_b)}}{N!} \right]
$$

$$
\times \left[\prod_{i=1}^{N} \frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b} \right] \tag{6.10}
$$

where N_b denotes the expected number of background events, N is the total number of observed events, σ is the systematic uncertainty in the estimate of the background, f_s and f_b are, respectively, the probability density functions for signal and background, and d_i are any kinematic parameters sensitive to the mass of the top quark (e.g., fitted mass, scalar sum of the transverse momenta of the 6 jets, etc.). The parameters m_t, n_s, and n_b denote, respectively, the mass of the top quark, the expected number of top-quark events, and the expected number of background events.

Because the uncertainty on the mass of the top quark is related directly to the sharpness of the log of the likelihood function near the maximum, it is important to choose selection criteria that provide as narrow a log-likelihood function as possible near that point. Taking the negative log of the likelihood function, we can write:

$$
l(m_t, n_s, n_b) \equiv -\log\left(\mathcal{L}(m_t, n_s, n_b)\right)
$$

$$= \log\left(\sqrt{2\pi}\sigma\right) + \left(n_b - N_b\right)^2/2\sigma^2 - N\log\left(n_s + n_b\right) + \left(n_s + n_b\right)$$

$$+ \log N! - \sum_{i=1}^{N}\log\frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b}. \tag{6.11}$$

Assuming that the negative of the log-likelihood function $l(m_t, n_s, n_b)$ has a local minimum at $(m_t^{min}, n_s^{min}, n_b^{min})$, and that it is approximately symmetric about the m_t, n_s and n_b axes at the minimum point, one could investigate the dependence on a change in the top mass by calculating the following quantity:

$$\delta l = l\left(m_t + \delta m_t, n_s^{min}, n_b^{min}\right) - l\left(m_t, n_s^{min}, n_b^{min}\right)$$

$$\approx \frac{\partial l\left(m_t, n_s^{min}, n_b^{min}\right)}{\partial m_t}\delta m_t. \tag{6.12}$$

Near the minimum, the quantity $\frac{\partial l(m_t, n_s, n_b)}{\partial m_t}\delta m_t$ can be rewritten as follows:

$$\frac{\partial l\left(m_t, n_s^{min}, n_b^{min}\right)}{\partial m_t} = \sum_{i=1}^{N}\frac{-1}{f_s(m_t, d_i) + n_b^{min}/n_s^{min}f_b(d_i)}\frac{\partial f_s(m_t, d_i)}{\partial m_t}. \tag{6.13}$$

Consequently, the dependence on the mass of the top quark (and thereby its uncertainty) can be changed by changing n_s, n_b and the probability density function for signal $f_s(m_t, d_i)$. One can therefore draw the following conclusions from Eq. (6.13):

- Because the quantity $\partial l(m_t, n_s, n_b)/\partial m_t$ is large for a narrow log-likelihood function, to reach the smallest uncertainty on the mass of the top quark one should keep $\partial l(m_t, n_s, n_b)/\partial m_t$ as large as possible by adjusting the kinematic acceptance criteria.

- Also, to reduce the uncertainty on the extracted mass, one should maximize the signal to background ratio to take advantage of the multiplicative term

$1/(f_s + n_b/n_s f_b)$.

- To minimize the uncertainty on the extracted mass, one should use the d_i and should find a kinematic acceptance criteria that provide a signal probability function $f_s(m_t, d_i)$ that is as sensitive as possible to the input mass m_t.

- Because n_s and n_b are constrained by the Poisson term in the likelihood function, f_s must be correlated with f_b. The impact of this correlation was studied previously [117] and discussed later in this chapter.

An important matter is therefore how to establish optimal selection criteria for the extraction of mass of the top quark that satisfy the above requirements.

6.2.1   Optimization

In order to minimize the statistical uncertainty on the mass of the top quark, an extension of the grid-search technique was used to establish selection criteria satisfying the requirements listed in the previous section. The specific procedure is described below.

After finding the optimal boundary that maximizes the signal to background, one calculates the difference between the average H_T values for the signal and for background at any given point on that boundary (ΔH_T) and uses that to improve mass extraction, as follows. Figure 6.11 (a) shows the optimal boundary that results from the full grid search. The grid points were sampled using the same data sets defined in the grid search described in the previous chapter. However, the grid search was performed using only H_T^{3j}, $\mathcal{C}$, and $\mathcal{A}$ in order to minimize degradation of the sensitivity on the input mass of the top quark.

The difference in mean H_T between signal and background is examined in Figs. 6.11 (b), (c) and (d). With the goal of obtaining selection criteria that would favor a signal probability density function $f_s(m_t, d)$ that is as different as possible from the background, we chose the points shown in Figs. 6.11 (b) and (c). This provided large ΔH_T, with a signal to background ratio that would still leave enough events to perform a mass analysis. Also, the selection favored a signal probability density function f_s that would be as sensitive as possible to the input top mass m_t by choosing a small H_T^{3j}, as in Fig. 6.11 (d). Ideally, one would calculate δl at every point in Fig. 6.11 (a) and then select the point that provides the largest δl. However, considering the available computing resources, that would indeed be a very challenging task to perform! In fact, what was done is that the H_T^{3j} cutoff was chosen to be as small as possible in order not to restrict unduly the range of allowed mass values. To summarize, the optimization procedure involved:

- Obtaining on the optimal boundary so as to make n_b/n_s to be small.

- Requiring H_T^{3j} to be small so as to make $\partial f_s(m_t, d_i)/\partial m_t$ large. (The evidence that this is a correct assessment, and that minimal requirements on H_T^{3j} provide greater sensitivity to mass, will be given in the next section.)

- Requiring ΔH_T large so as to make f_s differ as much as possible from f_b.

Arrows in Figs. 6.11 (a), (b), (c), and (d) show the operating point based on our qualitative optimization procedure, and Table 6.4 lists the selection criteria defined by the chosen point. Our strategy tends to reduce the uncertainty on the measurement of the mass of the top quark.
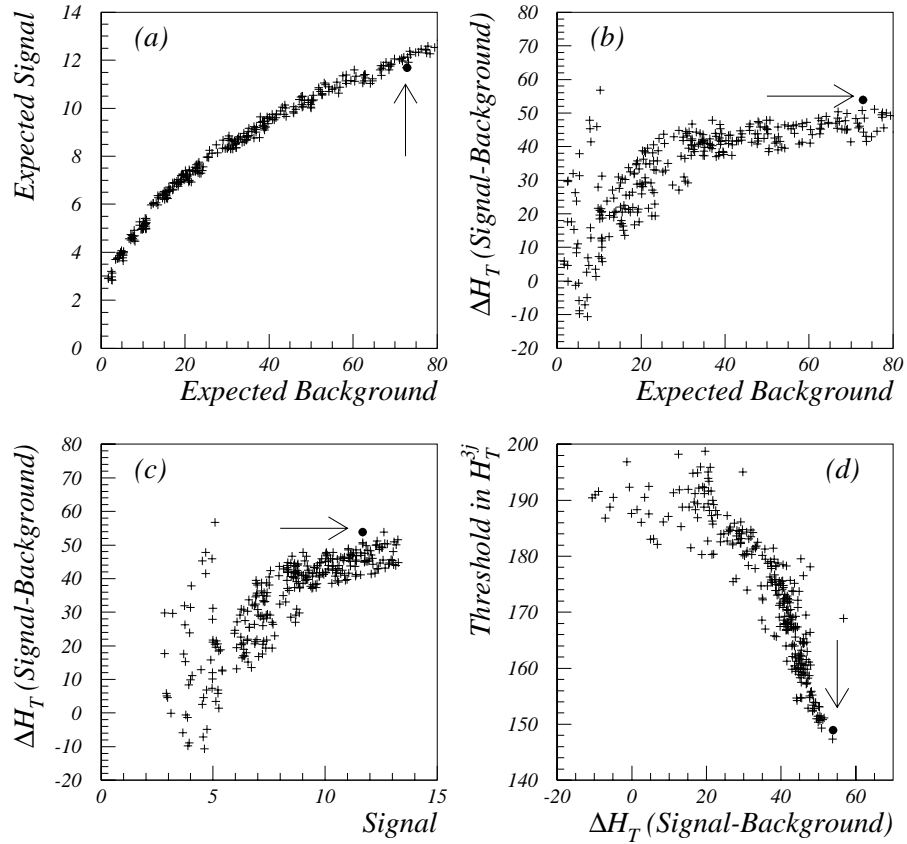
Figure 6.11: Results of the "mass-sensitive" Grid Search. Arrows and the dark circles indicate the point chosen for the mass optimization procedure. (a) Expected $t\bar{t}$ signal vs. background from previous grid search, with points shown only at the optimal boundary. (b) Difference between average H_T for signal and the background (ΔH_T) vs. expected background, (c) ΔH_T vs. expected signal, (d) H_T^{3j} vs. ΔH_T.

Table 6.4: Threshold selection criteria based on our "mass-sensitive" Grid Search.

variable	Centrality	Aplanarity	H_T^{3j} (GeV)
cutoff	> 0.68	> 0.061	> 149.0

6.2.2   Sensitivity

We repeated our investigation of the correlation between the mean fitted mass and the input mass, after applying the mass-sensitive selection criteria of Table 6.4. The relationship remains approximately linear (at the input mass points $m_t = 120$ and 140, statistics are poor), but the slope changes to 0.44 ± 0.017, and the offset to 110 ± 3.0 GeV/c^2, as shown in Fig. 6.12. The points were obtained using $t\bar{t}$ ISAJET for $m_t = 120$, 140, 160, 180, 200, and 220 GeV/c^2. The dot-dashed line is the result of a linear fit to the points, and the dotted line is along a 45° diagonal (reflecting perfect response). Also plotted (dashed line) is the result from Fig. 6.10, prior to imposing the additional cutoffs of Table 6.4. The increase in the offset value represents the impact of the mass sensitive selection criteria on the fitted mass. The slope, representing the effective sensitivity to the mass of the top quark, is degraded from 0.72 ± 0.014 to 0.44 ± 0.017.

Once we choose an operating point, the signal and background probability density functions, $f_s(m_t, d)$ and $f_b(d)$, can be constructed, and an estimate of the uncertainty expected on the value of the extracted mass can be made. The latter is obtained by generating an ensemble of $t\bar{t}$ events of some assumed mass, and defining an ensemble of background events. The value of the mass of the top quark is then extracted via the maximum likelihood fit, and the dispersion in these mass values, obtained from an ensemble of such "experiments", provides an estimate of the uncertainty in the extracted mass.

For the present study, as the mass sensitive discriminator (d), we chose the
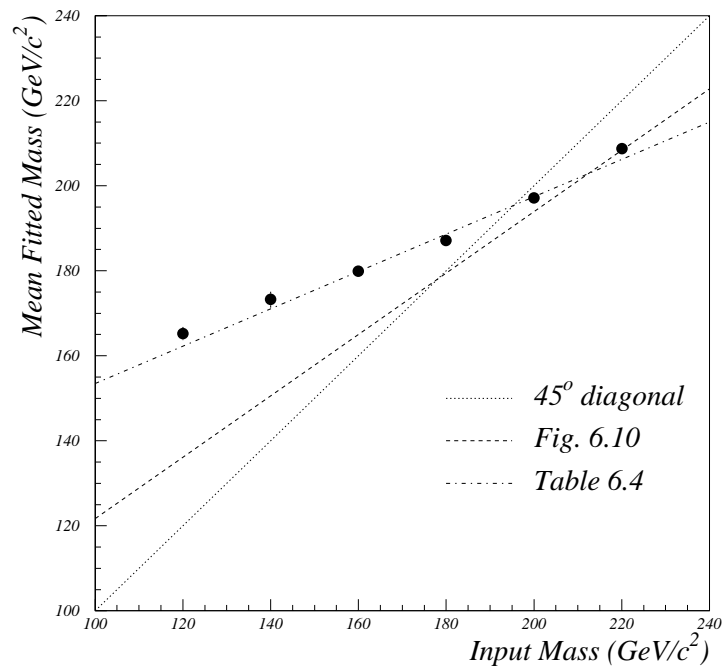
Figure 6.12: Mean fitted mass as a function of input mass, after imposing the mass-sensitive selection criteria in Table 6.4. The dashed line is the result of a linear fit given in Fig. 6.10, the dot-dashed line is the result of a linear fit to points, and dotted line along the diagonal is drawn for comparison. (The statistical errors are smaller than the size of the points.)
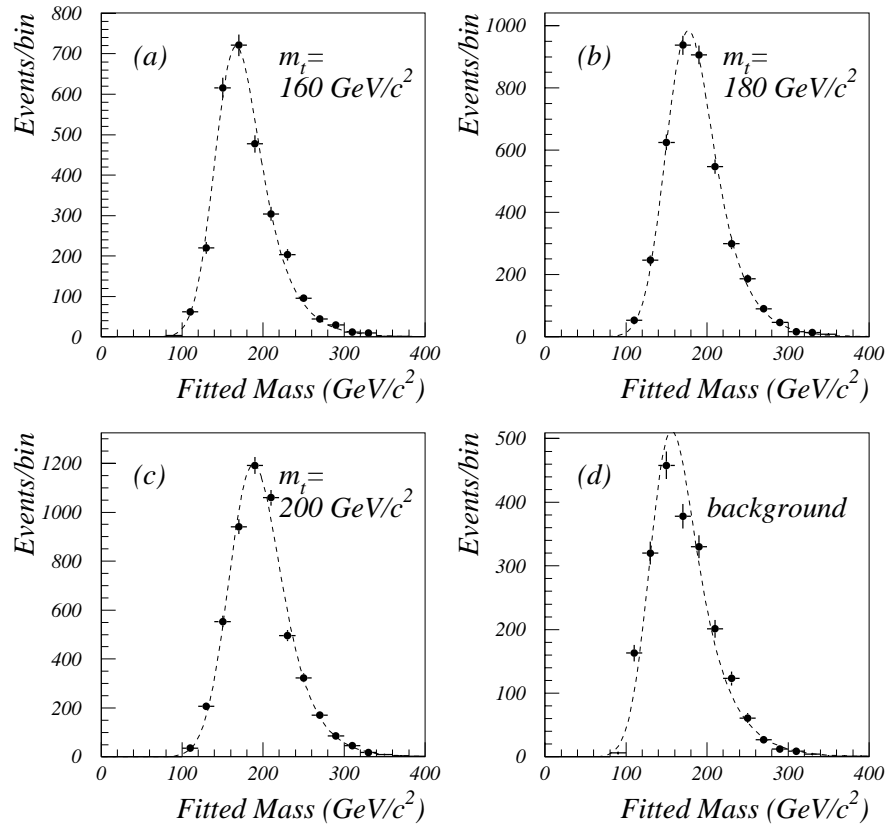
Figure 6.13: Distributions in m_t^{fit} for $t\bar{t}$ signal (using ISAJET) for three values of m_t, and background (data from RUN I), after application of our selection criteria of Table 6.4. Dashed curves are fits to the distributions using the functional forms given in the text. Results are for: (a) $m_t = 160$ GeV/c^2, (b) $m_t = 180$ GeV/c^2, (c) $m_t = 200$ GeV/c^2, and (d) background.

fitted mass (m^{fit}), based on the kinematics of the $t\bar{t}$ decay, as discussed in previous sections. The main advantage of doing a mass analysis with the fitted mass, as opposed to other parameters, is that m^{fit} should be the sharpest variable for extracting the mass of the top quark since, in principle, m^{fit} is the parameter that is most strongly correlated with the top mass. Using, for example, H_T for such analysis[118] is more likely to yield an incorrect mass of the top quark, particularly if $t\bar{t}$ production has a non-negligible contribution from non-standard sources (e.g., heavy $t\bar{t}$ resonances, as suggested by Hill and Parke[119], etc.).

Prior to performing the kinematic fitting, the standard DØ jet-energy correction package (CAFIX) as well as the OOC-correction were applied to jets in the search and the $t\bar{t}$ Monte Carlo samples. Then, a function with a multiplicative exponential decay term was adopted as the form to model the probability density for background and signal. In particular, the probability function used to characterize the signal was:

$$f_s(m_t, d) = N G((d - p_1)/p_2) e^{-p_3 d} \qquad (6.14)$$

where

$$
\begin{aligned}
p_1 &= a_1 + a_2 m_t + a_3 (m_t)^2 \qquad &(6.15)\\
p_2 &= a_4 + a_5 m_t + a_6 (m_t)^2 \\
p_3 &= a_7 + a_8 m_t \\
d &= m^{fit},
\end{aligned}
$$

N is a normalization constant, a_i are free parameters in the fit, and $G(x)$ is the frequency function, defined as

$$G(\frac{d - p_1}{p_2}) \equiv G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt. \qquad (6.16)$$

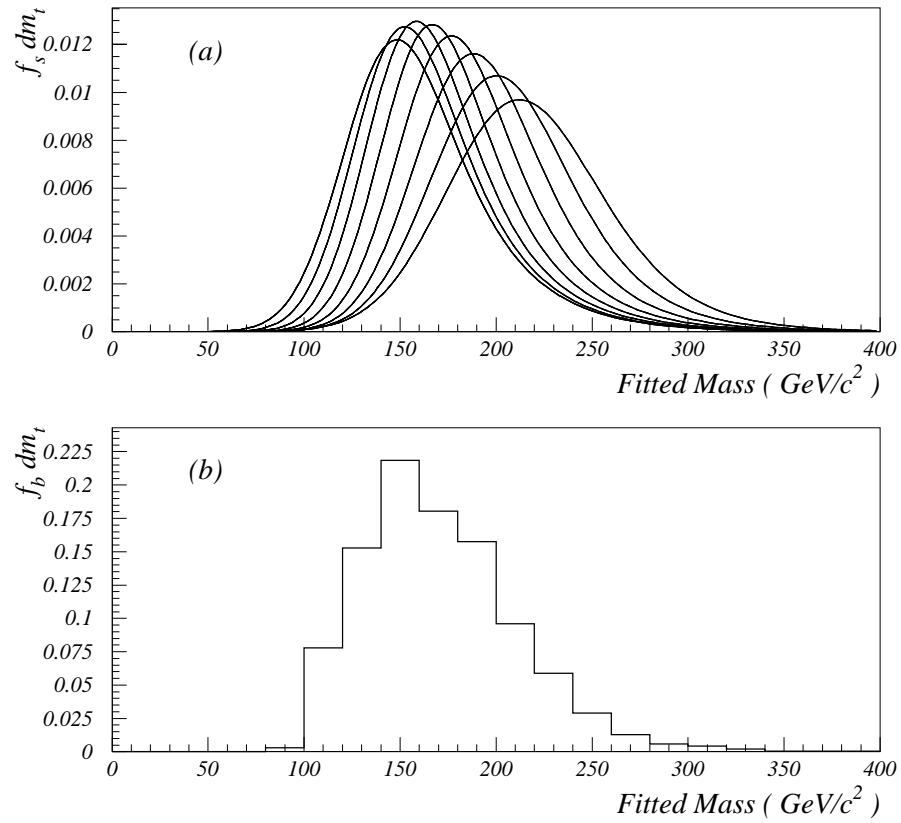Figure 6.14: Probability density functions for signal and background after application of the selection criteria of Table 6.4. (a) Signal probability density functions are for a top-quark mass of 100 to 240 GeV/c^2 in mass intervals of Δm_t=20 GeV/c^2. (b) Probability density function (histogram) for the background (same as Fig. 6.13 (d)).

The probability density function for the background uses the same functional form where d is some mass parameter, but there are no m_t-dependent terms in the p_i. Therefore, the probability function used to characterize the background became also:

$$f_b(d) = NG((d - p_1)/p_2)e^{-p_3 d} \qquad (6.17)$$

Figures 6.13 (a), (b), and (c) show the distributions in fitted mass for three fixed input masses as the result of a two-dimensional fit. The signal probability functions were generated using ISAJET. Figure 6.13 (d) shows the results of the fit to background. The fit to background was very poor and was not improved by using other functional form probably because of the peculiar shape of the distribution near the peak. Therefore, in what follows, we will use the histogram itself (points in Figs. 6.13 (d)) as our primary choice of the probability density function for background. The quality of the fits to signal will be discussed later when we turn to systematic uncertainties. Figure 6.14 (a) shows the probability density function $f_s(m_t, m^{fit})$ for different input top masses in 20 GeV/c^2 intervals, and (b) shows, for comparison, the background probability density function $f_b(m^{fit})$ given in Fig. 6.13 (d). The background shape is similar to the signal for an input top-quark mass of $m_t \approx 140$ GeV/c^2.

Because the background was taken as the data from RUN I, we expect it to contain some contribution from top production, which can may bias the *shape* of the background. To minimize the effect of this bias, we defined the following corrected background function:

$$f_b = \frac{1}{1 - C}(f_b - C f_s) \qquad (6.18)$$

where C is the fraction of events in the background sample that can be attributed

Table 6.5: Relationship between the signal to background ratio and the expected uncertainty in extracted mass, as obtained from ensemble tests.

signal / background	RMS (GeV/c^2)	Expectation from $\sqrt{n_s}$ scaling (GeV/c^2)
100/0	8.0 ± 0.13	8.0
70/30	11.1 ± 0.18	9.5
50/50	17.1 ± 0.27	11.3
30/70	30.1 ± 0.48	14.6

to signal ($t\bar{t}$ production) at any given mass m_t. The value of C was estimated from the value of n_s that is iterated during the likelihood fit procedure.

The effect of the signal to background ratio on the uncertainty in the mass was discussed in the previous section. To study this in more detail, we generated ensembles of 100 events, assuming different signal to background ratios, for an input top-quark mass of $m_t = 180$ GeV/c^2. In order to quantify directly the dependence of the uncertainty on the signal to background ratio, we do not allow the total number of observed events to fluctuate. Figures 6.15 (a), (b), (c), and (d) show the distributions of the maximum likelihood mass estimates for different signal to background ratios: 100/0, 70/30, 50/50, and 30/70, respectively. These plots confirm that our entire maximum likelihood procedure is sound. Table 6.5 lists the standard deviations (RMS) for the four sets of distributions together with expectations from the $\sqrt{n_s}$ scaling rule. It is clear from these results that the uncertainty on the top quark mass is larger for smaller values of the signal to background ratio, beyond that expected just from the $\sqrt{n_s}$ scaling rule.

To estimate the uncertainty in the extracted mass for the selection criteria given in Table 6.4, we generated ensembles of events assuming the expected number of signal and background events. (The procedure for choosing the number of signal and background events will be discussed shortly.) We allow the total number of
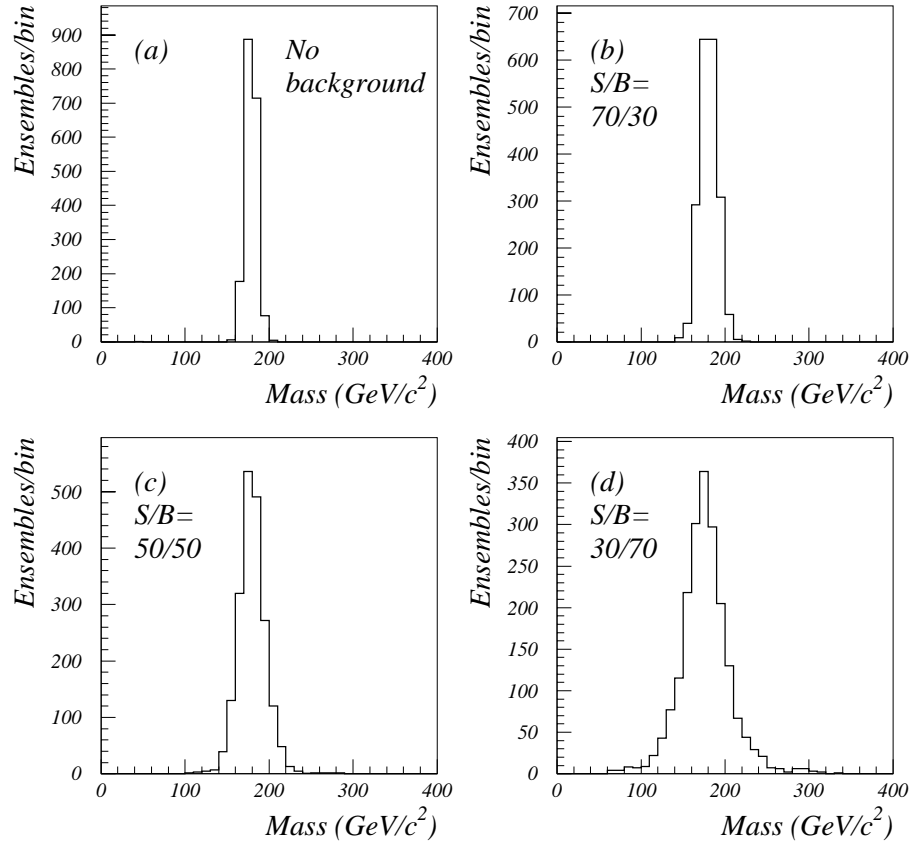
Figure 6.15: Results from ensemble tests of the likelihood fit using 2000 samples of 100 events. Signal events are from $t\bar{t}$ ISAJET at $m_t = 180$ GeV/c^2. Background events are from the search sample. The extracted mass values are for (a) signal to background ratio of S/B=100/0, (b) S/B=70/30, (c) S/B=50/50, and (d) S/B=30/70.

Table 6.6: Relationship between mass of the top quark and the expected uncertainty in the extracted mass, from simulations of ensembles that have event samples comparable to the data from RUN I.

input m_t (GeV/c^2)	RMS (GeV/c^2)
140	29.6 ± 0.63
160	29.2 ± 0.57
180	27.7 ± 0.50
200	24.4 ± 0.52

observed events to fluctuate according to Poisson statistics, which provides a more realistic estimate of the uncertainty in the mass of the top quark. Figures 6.16 (a), (b), (c) and (d) show the distribution of extracted mass values that minimize the negative log-likelihood function for ensembles of 5000 events, generated with ISAJET for input top-quark masses of $m_t = 140$, 160, 180, and 200 GeV/c^2, respectively. The background events were chosen at random from data (search sample) of RUN I. All events were required to pass the kinematic criteria given in Table 6.4. The RMS values of the Gaussian fits to distributions, which are given in Table 6.6 as a function of the assumed top quark mass, reflect the expected uncertainty in the extracted mass values.

6.2.3 Likelihood Fit with Data

The search sample that excluded the 7219 events discussed in the previous chapter is again used for mass extraction. A total of 2087 events remain after imposing the selection criteria and kinematic fitting, and 83 events after imposing b-jet tagging. The remaining background, corrected using the prescription given in the previous chapter (Eq. (5.27)), was estimated as 61.3 events. Table 6.7 summarizes these numbers. An "unbinned" likelihood fit was performed on the resulting data sample, which gave a top quark mass estimate of $m_t = 153.6$ GeV/c^2.
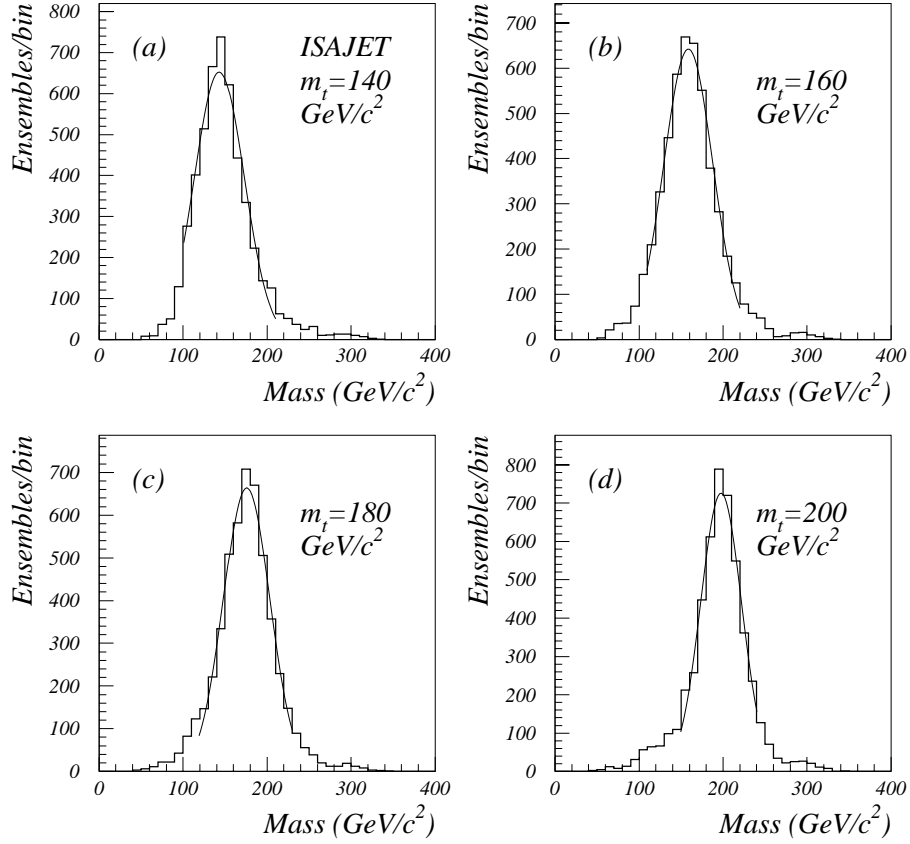
Figure 6.16: Results of likelihood fits to event samples consisting of N_{obs}=83 and N_{bkg}=61 events, for $t\bar{t}$ ISAJET for (a) m_t=140, (b) m_t=160, (c) m_t=180, and (d) m_t=200 GeV/c^2.

Table 6.7: Result of the imposition of criteria of Table 6.4 on data from RUN I.

Selection criteria	Number of events
Kinematic criteria	2087
Above, with muon tagging	83
Expected background	61.3
Systematic uncertainty in background	7.4

An approximation to the standard deviation on the mass can be obtained from the likelihood curve from a change in the value of $-\log(\mathcal{L})$ by 0.5 unit. This procedure yields a statistical uncertainty of $^{+35.2}_{-37.5}$ GeV/c^2. From Table 6.6, the standard deviation expected on the basis of ensemble studies is ~ 29 GeV/c^2. Because getting the uncertainty from the likelihood curve can be unreliable, we chose to use the uncertainty obtained from the ensemble tests that is less subject to fluctuations.

Figure 6.17 (a) shows contours of $-\log(\mathcal{L}) = 0.25$, 0.5, and 1.0 as a function of n_s and mass, indicating some correlation between the two parameters. Figure 6.17 (b) shows the same contours in n_b vs. mass space, again showing some correlation. Figure 6.17 (c) shows the minimum value of $-\log(\mathcal{L})$ vs. mass (n_s and n_b are allowed to vary to find a minimum at each mass value. This is equivalent to performing MINOS minimization in MINUIT[116]). Figure 6.17 (d) shows the distributions in fitted mass (black points), with the dark-shaded area representing the signal and light-shaded area representing the background (histogram), obtained from the likelihood fit. Table 6.8 summarizes the values of the parameters obtained in the likelihood fit, where the quoted uncertainties correspond to change of 0.5 in $-\log(\mathcal{L})$. The large uncertainty on n_s indicates that the current analysis does not provide a significant excess of signal to support the presence of a $t\bar{t}$ contribution to multi-jet final states.

As a check of our procedure, the likelihood fit was redone without the Gaussian constraint on n_b. The value of n_b obtained from this fit is then a background estimate that is less dependent on results obtained in the previous chapter, but dependent more on the shape of the mass distributions. The likelihood fit gave a mass of 149.2 $^{+51.9}_{-28.6}$ GeV/c^2, $n_s = 36.6$ $^{+29.9}_{-33.6}$, and $n_b = 46.4$ $^{+34.7}_{-28.8}$. This again suggests that there is no strong evidence for $t\bar{t}$ signal in the data.

Table 6.8: Result of the likelihood fit to data from RUN I.

	fit result
m_t (GeV/c^2)	$153.6 \; ^{+35.2}_{-37.5}$
n_s	$23.1 \; ^{+11.5}_{-11.0}$
n_b	$60.6 \; ^{+7.2}_{-7.2}$
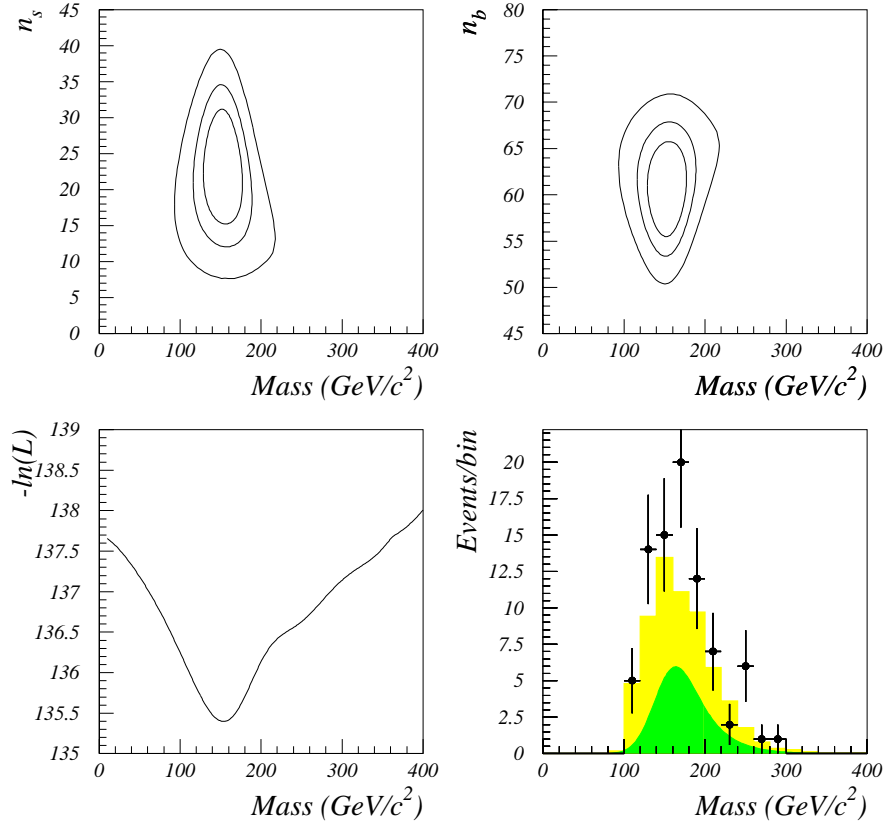
Figure 6.17: Results from the likelihood fit to the data. (a) Contours of $-\log(\mathcal{L}) = 0.25$, 0.5, and 1.0 as a function of n_s and mass, (b) contours of $-\log(\mathcal{L}) = 0.25$, 0.5, and 1.0, as a function of n_b and mass, (c) $-\log(\mathcal{L})$ as a function of mass, (d) distribution of m_i^{fit} for our sample of 83 events (points), with the dark-shaded area representing the fitted signal, and light-shaded histogram representing the background.

Table 6.9: Relation between the input top-quark mass and the expected shift due to a bias in the analysis procedure.

input m_t (GeV/c^2)	observed shift (GeV/c^2)
140	$+ 9.6 \pm 0.52$
160	$+ 1.1 \pm 0.51$
180	$- 4.8 \pm 0.51$
200	$- 5.7 \pm 0.51$

6.2.4   Systematic Uncertainties

The following comprise the major sources of systematic uncertainty that affect the extraction of the mass of the top quark in this analysis.

1. Shift of extracted mass from the input top mass due to a bias in the procedure: Figure 6.16 shows a shift in the extracted mass that depends on the input mass. The systematic shift is defined as the difference between the input masses and the average of the values obtained in the likelihood fits. Table 6.9 gives this shift as a function of the input mass. Of course, this dependence can be used to correct the final result, but the uncertainty in the shift contributes to the overall uncertainty in the extracted mass.

2. Jet energy scale: Uncertainty on the mass of the top quark due to the uncertainty on the jet energy scale was estimated as follows. Assuming the current uncertainty on energy of $\pm$ 5 %, ensembles of events were reprocessed using jet energies scaled up by 1.05 and down by 0.95. The resulting fitted mass distributions for the signal and background are shown in Fig. 6.18. We see that the $\pm$ 5 % change in jet energy scale changes the mean fitted mass by approximately $\pm$ 7.5 GeV/c^2, regardless of input mass. A similar shift ($\pm$ 6.7 GeV/c^2) was found for the background distribution. Figures 6.19 (a), (b), (c), and (d) show the mass distributions resulting from the likelihood

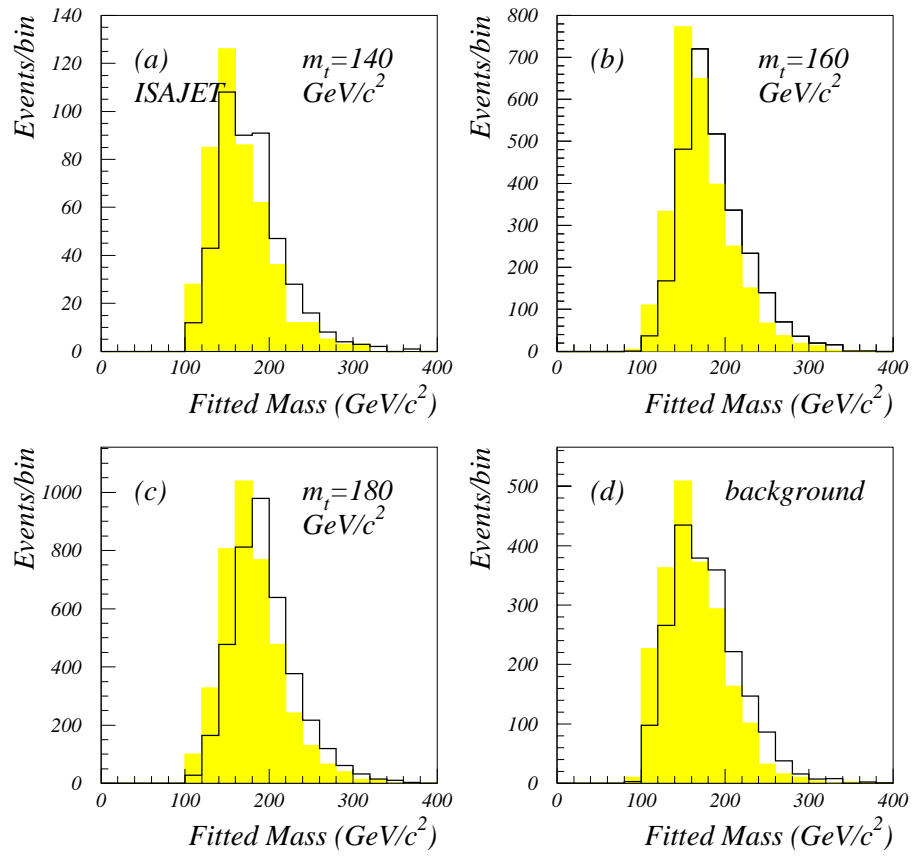Figure 6.18: Results from fitting of $t\bar{t}$ signal and background events when jet energies were scaled up by 5 %(normal histograms), and scaled down by 5 %(shaded histograms) for: (a) m_t=140 GeV/c^2, (b) m_t=160 GeV/c^2, (c) m_t=180 GeV/c^2, and (d) background.

Table 6.10: Relation between the input top-quark mass and the expected uncertainty due to a $\pm$ 5 % uncertainty in the jet-energy scale.

input m_t (GeV/c^2)	140	160	180	200
(+) uncertainty (GeV/c^2)	+48.6	+37.1	+25.7	+18.3
(−) uncertainty (GeV/c^2)	−25.4	−38.8	−47.0	−52.3

fit for different input masses. The systematic uncertainty in the mass is taken as the mean of the offsets between the result of the input mass and the energy-shifted distributions. Table 6.10 gives the uncertainty as a function of the input top-quark mass.

3. Models of parton fragmentation : In this analysis, we used primarily the ISAJET event generator. By repeating the fitting with HERWIG samples of m_t = 160, 180, and 200 GeV/c^2, we set an expected upper limit on changes in mean values of fitted mass distribution as − 6 GeV/c^2 (HERWIG preferred the lower value of the fitted mass). Based on Fig. 6.12, we estimated an approximation to the uncertainty due to different models of parton fragmentation as −13.0 GeV/c^2.

Adding the systematic uncertainties in quadrature (for an input top-quark mass 150 GeV/c^2), we obtain a corrected mass of the top quark of 148±29 (statistical) $^{+43}_{-35}$ (systematic) GeV/c^2.

6.3 Approach using Bayesian Theory

We also attempted an alternative way to estimate the mass of the top quark using Bayesian probability theory[117]. In the top-quark mass analysis, the desired quantity is not the likelihood $P(D|m_t, I)$, but rather the a posteriori probability
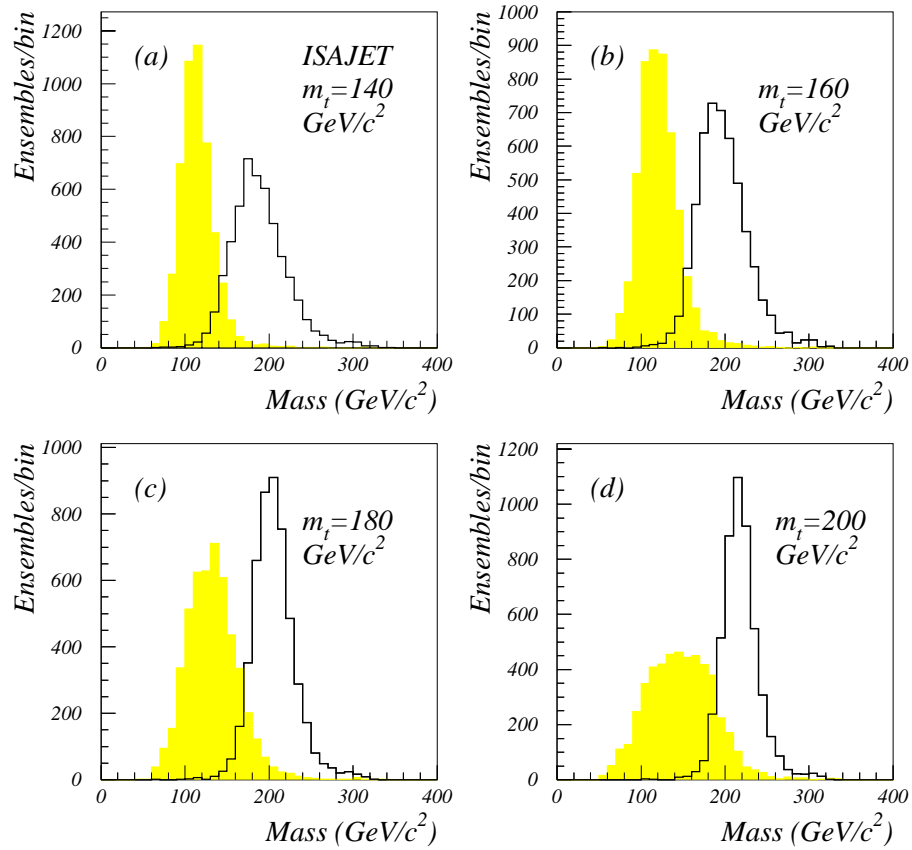
Figure 6.19: Results from ensemble tests of N_{obs}=83 and N_{bkg}=61 events, when jet energies were scaled up by 5 %(normal histograms), and scaled down by 5 %(shaded histograms) for; (a) m_t=140 GeV/c^2, (b) m_t=160 GeV/c^2, (c) m_t=180 GeV/c^2, and (d) m_t=200 GeV/c^2.

$P(m_t | D, I)$, that is, the a posteriori probability of the top-quark mass being m_t, given D, the observations or experimental facts and I, any a priori assumptions. In case n_s and n_b are also of interest, Bayes' theorem provides:

$$P(m_t, n_s, n_b | D, I) = \frac{P(D | m_t, n_s, n_b, I) P(m_t, n_s, n_b | I)}{\int dm_t dn_s dn_b P(D | m_t, n_s, n_b, I) P(m_t, n_s, n_b | I)} \qquad (6.19)$$

and the a posteriori probability $P(m_t | D, I)$ is obtained by simply marginalizing Eq. (6.19) with respect to n_s and n_b, that is, by performing integrations over n_s and n_b, as we discussed in the previous chapter.

CHOICE OF THE LIKELIHOOD

For our likelihood we can use the following:

$$P(D | m_t, n_s, n_b, I) = \prod_{i=1}^{N} \frac{e^{-n} n^{N_i}}{N_i!} \qquad (6.20)$$

where

$$n(m_t, n_s, n_b) = n_s f_s^i(m_t) + n_b f_b^i, \qquad (6.21)$$

N_i is the number of events in the ith bin, N is the total number of bins, f_s^i and f_b^i are the values of signal and background probability density functions at the ith bin, respectively, and n_s and n_b are the number of signal and background expected, respectively. Thus, Eq. (6.20) is, in fact, a standard binned likelihood function. Another choice for the likelihood can be one similar to that used earlier (see Section 6.2):

$$
\begin{aligned}
P(D | m_t, n_s, n_b, I) &= \left[\frac{(n_s + n_b)^N e^{-(n_s + n_b)}}{N!} \right] \\
&\times \left[\prod_{i=1}^{N} \frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b} \right].
\end{aligned} \qquad (6.22)
$$

But, in general, such an unbinned likelihood function always has a term that is not a function of m_t. This can cause divergences for integration of $P(m_t|D,I)$[120]. That is, it is possible to have the constant term sizable enough that the integral $\int dm_t m_t P(m_t|D,I)$ (mean estimate on m_t) becomes not well defined.

CHOICE OF THE A PRIORI PROBABILITY

The a priori probability $P(m_t, n_s, n_b|I)$ must, somehow, express total ignorance of the values of m_t, n_s, and n_b. Whether this is possible, or how best to do that, is an unresolved issue. One choice is the uniform a priori probability:

$$P(m_t, n_s, n_b|I) \times dm_t dn_s dn_b = constant \times dm_t dn_s dn_b. \tag{6.23}$$

The problem with a uniform prior is that it leads a logical inconsistency that (1=2) $< n > = n + 1$ for Poisson distributions[104]. Another choice is Jeffrey's prior[97], which for the mass of the top quark can be written as:

$$P(m_t, n_s, n_b|I) = \frac{1}{m_t}. \tag{6.24}$$

The number of background events N_b and the uncertainty σ on N_b, can be taken as a priori knowledge of n_b:

$$P(m_t, n_s, n_b|I) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(n_b - N_b)^2/2\sigma^2}. \tag{6.25}$$

It has been argued, however, that there may be no such thing as a non-informative a priori probability[104, 120]. In that case, the Bayesian approach will not provide the notion of an absolutely "objective" method of analysis. Nevertheless, what really matters, in the Bayesian approach, is how one makes best inferences on parameters based on what is actually observed (data) and on what

one knows regardless of the actual observation (a priori information).

6.3.1 Monte Carlo Tests

To compare the Bayesian and traditional likelihood methods, the selection criteria of Table 6.4 were also used in this analysis. However, a signal to background ratio of 50 to 50 was chosen to test the efficacy of the Bayesian formulation. Using the unbinned likelihood given in Eq. (6.22), with an m_t-independent (uniform) prior for m_t, and with a Gaussian prior for n_b, we calculated the a posteriori probabilities. After the imposition of the selection criteria of Table 6.4, a sample of fifty background events and fifty signal were retained from data of RUN I and from ISAJET $t\bar{t}$ production at m_t=180 GeV/c^2, respectively. Figure 6.20 (a) shows the a posteriori probability, $P(m_t, n_s | D, I)$, as a function of the number of top events, and of the mass of the top quark, indicating the expected signal. Figure 6.20 (b) shows contours of equal probability. The distributions show a peak near the input values of m_t and n_s, demonstrating the clear signal in (n_s, m_t) space. Figure 6.20 (c) shows the a posteriori probability distribution $P(m_t | D, I)$. The mean estimate for m_t is 184.9 GeV/c^2, and, more importantly, the RMS uncertainty on m_t is 18.7 GeV/c^2, which can be compared with the result from the likelihood fit of 17.1 GeV/c^2 in Fig. 6.15 (c) (or in Table 6.5). It can be argued that because the uncertainty from the likelihood fit was based on fits to each ensemble of many pseudo experiments, it is not fair to compare the two numbers. However, a fundamental notion in Bayesian statistics is that inferences should be based on what is actually observed, and not on what we could have observed. Therefore, regardless of the fairness of the comparison, the sampling distribution of Bayesian estimates cannot alter the inferences based on some given data set. This is because the a posteriori probability, by construction, incorporates the likelihood (and therefore the actually

observed data) and the assumed a priori probability, and, necessarily, describes all that we know about the hypothesis being tested. This is, precisely the importance of Bayes' theorem.

Since $P(m_t|D, I)$ is symmetric, and its shape is approximately Gaussian, one expects that the square root of the variance and the 68.3% confidence interval estimate will provide similar answers (as should the mean and the median estimates in this case). In general, however, these uncertainty measures can be different.

For our second choice of the likelihood, we used the binned likelihood function given in Eq. (6.20). The same signal to background ratio of 50 to 50 was used again for our study. We again assumed a Gaussian prior for n_b and an m_t-independent (uniform) prior for m_t, and obtained a mean estimate on m_t of 182.7 GeV/c^2 and an RMS uncertainty on m_t of 19.1 GeV/c^2, close to the result using an unbinned likelihood. Figures 6.21 (a), (b), and (c) show the a posteriori probabilities for the binned likelihoods, which are quite similar to those in Fig. 6.20.

When we assume a $1/m_t$ form for the prior on the mass of the top quark, and Gaussian prior for n_b, with the unbinned likelihood, the mean and the RMS uncertainty on the top-quark mass become 183.0 GeV/c^2 and 18.8 GeV/c^2, respectively, and with the binned likelihood the mean and the RMS uncertainty are 180.7 GeV/c^2 and 19.1 GeV/c^2, respectively.

As expected, when $1/m_t$ is assumed for the a priori probability and flat a priori probability for n_b, the mean value of m_t becomes lower than for the case of a uniform prior, because $1/m_t$ favors lower values of top-quark mass. However, for a uniform prior on n_b, the procedure deteriorates badly. Figure 6.22 (a) and (b) show the posterior probabilities with unbinned and binned likelihoods, respectively. In both cases, the a posteriori probabilities have peaks at zero which may be due to the correlation between the $1/m_t$ prior on m_t and what is used for n_b. Clearly, the most probable values for the uniform priors on n_b and the $1/m_t$ priors on m_t, reflect
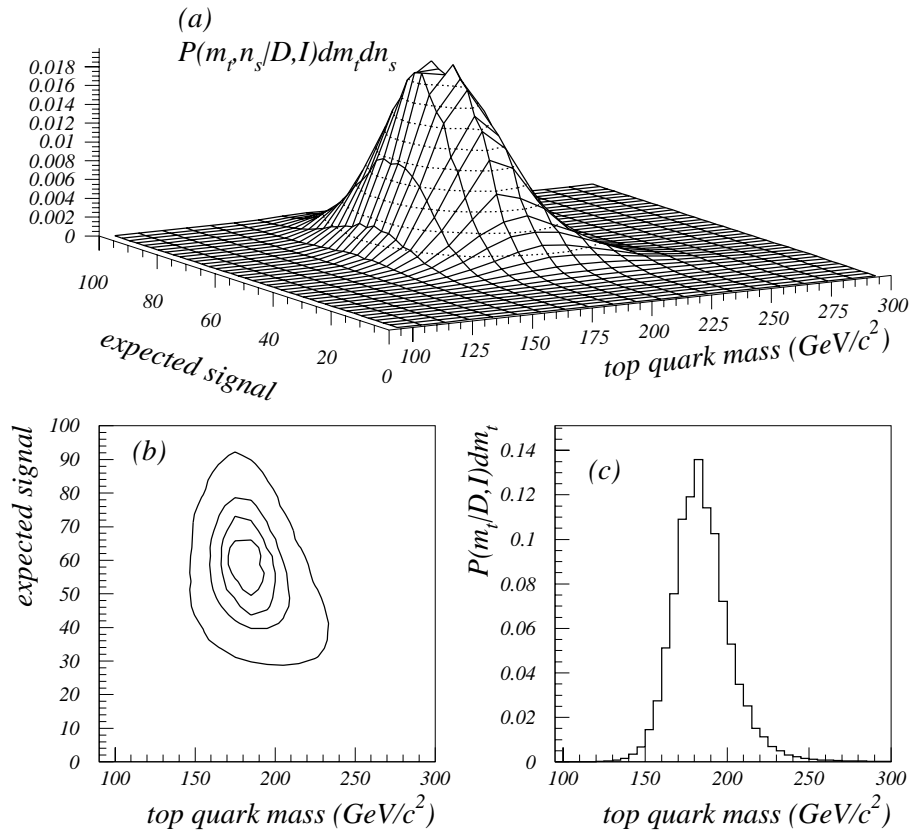
Figure 6.20: A posteriori probability distributions using an unbinned likelihood, a uniform prior for m_t and a Gaussian prior for n_b. A total of 50 signal events were chosen from $t\bar{t}$ ISAJET at $m_t = 180$ GeV/c^2, and 50 background events from data of RUN I. (a) $P(m_t, n_s | D, I)$ as a function of m_t and n_s. (b) Contours of equal probabilities (0.016, 0.011, 0.006, and 0.001) for (a). (c) $P(m_t | D, I)$ as a function of m_t.

Figure 6.21: A posteriori probability distributions using a binned likelihood, a uniform prior for m_t and a Gaussian prior for n_b. A total of 50 signal events were chosen from $t\bar{t}$ ISAJET at $m_t = 180$ GeV/c^2, and 50 background events from data of RUN I. (a) $P(m_t, n_s|D, I)$ as a function of m_t and n_s. (b) Contours of equal probabilities (0.016, 0.011, 0.006 and 0.001) for (a). (c) $P(m_t|D, I)$ as a function of m_t.
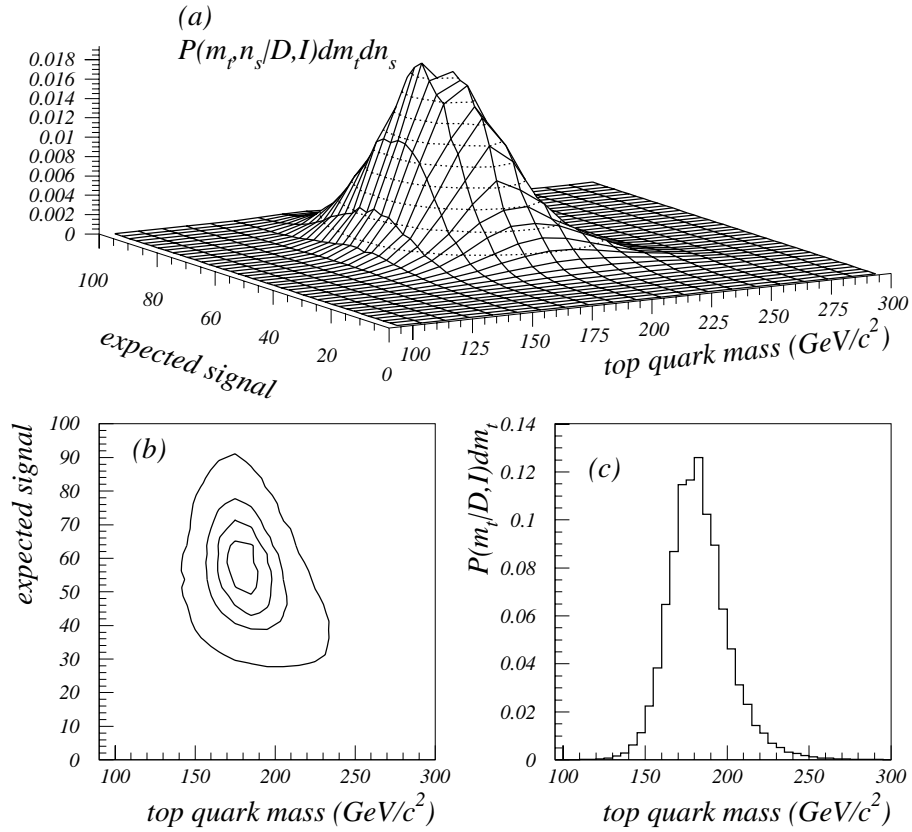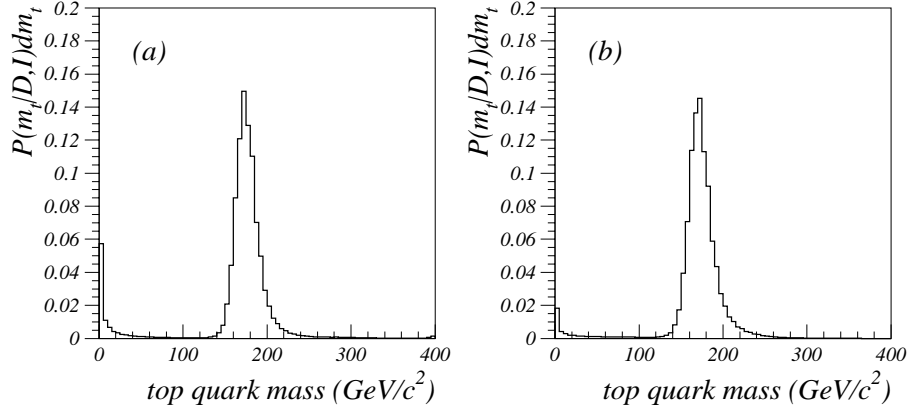
Figure 6.22: A posteriori probability distributions using $1/m_t$ as a priori for m_t, and a uniform prior on n_b. A total of 50 signal events were chosen from $t\bar{t}$ ISAJET at $m_t = 180$ GeV/c^2, and 50 background events from data of RUN I. (a) $P(m_t|D,I)$ as a function of m_t using unbinned likelihood, (b) $P(m_t|D,I)$ as a function of m_t using binned likelihood.

the input mass much better than do the means of the a posteriori probabilities. Table 6.11 summarized the results using different combinations of likelihoods and a priori probabilities.

We performed the same analysis for an input mass of 140 GeV/c^2. For the unbinned likelihood, we obtained a mean estimate of 133.7 GeV/c^2 and the RMS uncertainty of 21.8 GeV/c^2. With the binned likelihood, we obtained the mean estimate as 135.5 GeV/c^2 and the uncertainty of 24.5 GeV/c^2. In both cases, we assumed Gaussian a priori probability for n_b and a flat prior for m_t.

Our results indicate that imposing $1/m_t$ and an uniform prior for n_b can introduce a bias in the measurement. However, it should be recognized, again, that the concept of "bias" is not critical to Bayesian inference, because what matters is that the Bayesian estimate be "consistent", that is, converge to the true value of mass as more data are accumulated.

Table 6.11: The mean, RMS, mode, and uncertainty on the top mass for different combinations of likelihoods and priors. An input mass of 180 GeV/c^2 and a signal to background ratio of 50/50 were assumed in these Monte Carlo studies of the Bayesian approach.

	mean $\pm$ RMS (GeV/c^2)	mode $\pm$ one s.d. (GeV/c^2)
unbinned $P(D\|m_t, n_s, n_b, I)$ uniform in n_b uniform in m_t	181.1 ± 30.9	175.0 ± 12.2
unbinned $P(D\|m_t, n_s, n_b, I)$ Gaussian in n_b uniform in m_t	184.9 ± 18.7	182.0 ± 14.4
unbinned $P(D\|m_t, n_s, n_b, I)$ uniform in n_b $1/m_t$ in m_t	162.8 ± 54.6	174.2 ± 12.1
unbinned $P(D\|m_t, n_s, n_b, I)$ Gaussian in n_b $1/m_t$ in m_t	183.0 ± 18.8	180.8 ± 14.3
binned $P(D\|m_t, n_s, n_b, I)$ uniform in n_b uniform in m_t	177.1 ± 23.9	171.7 ± 13.6
binned $P(D\|m_t, n_s, n_b, I)$ Gaussian in n_b uniform in m_t	182.7 ± 19.1	179.5 ± 15.1
binned $P(D\|m_t, n_s, n_b, I)$ uniform in n_b $1/m_t$ in m_t	168.9 ± 37.3	170.3 ± 13.8
binned $P(D\|m_t, n_s, n_b, I)$ Gaussian in n_b $1/m_t$ in m_t	180.7 ± 19.1	178.4 ± 15.2

6.3.2 Extraction of the Mass of the Top Quark using Bayesian Theory

Data used in the likelihood-fit analysis were used again to obtain a Bayesian estimate of the mass of the top quark. All the event selection criteria, including jet energy corrections, were done in exactly the same way as in the previous likelihood-fit analysis. Consequently, the results given in Table 6.7 are still valid for this analysis. To compare directly with the results of the likelihood fit, we take an unbinned likelihood, uniform prior for m_t, and a Gaussian prior for n_b. The resultant a posteriori probability $P(m_t, n_s | D, I)$ is shown in the Figs. 6.23 (a) and (b). There is a small rise of the probability at very high m_t, near 400 GeV/c^2, where n_s is near zero. We should note that above $m_t = 220$ GeV/c^2 the signal probability density function f_s is extrapolated from a two-dimensional fit, and it may therefore not be valid for masses beyond $m_t = 220$ GeV/c^2. There is another, more significant rise centered near $m_t \approx 150$ GeV/c^2 that is correlated with the value of n_s that can possibly be explained by the small excess of events near $m^{fit} \approx 180$ GeV/c^2 observed in Fig. 6.17 (d). Figure 6.23 (c) shows the a posteriori probability, $P(n_s | D, I)$ as a function of the expected signal, n_s. The mean estimate of n_s and its RMS uncertainty is 19.3 ± 11.0, which is comparable to the values obtained in the likelihood-fit analysis, which is $23.1 \, ^{+11.5}_{-11.0}$. Figure 6.23 (d) shows the a posteriori probability $P(m_t | D, I)$ as a function of m_t. A peak in the distribution is observed at 149.2, with an one standard deviation uncertainty (from a Gaussian fit to the peak) of 37.3 GeV/c^2. This is also comparable with the result from the likelihood-fit analysis, which is $153.6 \, ^{+35.2}_{-37.5}$ GeV/c^2. One philosophical difference between the present treatment and the standard approach should be noted, and that is that here the parameters n_b and n_s have been integrated out, whereas in the likelihood fit only one value of n_b and n_s, namely n_b^{min} and n_s^{min},

were considered in the extraction of the central value of m_t.

6.3.3 Systematic Uncertainty

The jet energy scale appears to be a primary source of systematic uncertainty in the extraction of the mass of the top-quark. In what follows, we treat this uncertainty as one more piece of a priori information that is provided to us. Suppose that the fractional uncertainty on the fitted mass is q,[*] we can then write:

$$m^{fit}(q) = (1 + q)m^{fit}_{measured} \qquad (6.26)$$

where $m^{fit}(q)$ denotes the rescaled fitted mass and $m^{fit}_{measured}$ the original fitted mass, respectively. Because rescaling the fitted mass changes the shape of signal and background as well as the distribution of events, f_s and f_b must therefore also depend on q:

$$f_s(m_t, d_i) \rightarrow f_s(m_t, d_i(q))$$
$$f_b(d_i) \rightarrow f_b(d_i(q))$$

The a posteriori probability that takes proper account of the uncertainty in the jet energy scale is given by the following integral:

$$
\begin{aligned}
P(m_t|D, I) &= \int_{q_{low}}^{q_{high}} dq P(m_t|D, I) \\
&= \frac{1}{N} \int_{q_{low}}^{q_{high}} dq P(D|m_t, q, I) P(q|I) P(m_t|I), \qquad (6.27)
\end{aligned}
$$

---

[*]Ideally, one should assume $E_{jet}(q) = (1 + q) E^{jet}_{measured}$, and repeat the kinematic fitting procedure, rather than rescale the fitted mass values directly. However, repeating the fitting in each steps of the integration in order to obtain probability density functions is too time consuming, and we will therefore assume a uniform uncertainty on the fitted mass as 7.5 % due to the jet-scale uncertainty, based on our earlier study (see Section 6.2.4).

Figure 6.23: A posteriori probability distributions using an unbinned likelihood, a uniform prior for m_t and Gaussian prior for n_b. The data are the 83 events from RUN I analyzed previously using the standard kinematic fitting procedure. (a) $P(m_t, n_s | D, I)$ as a function of m_t and n_s. (b) Contours of equal probability for (a), (c) $P(n_s | D, I)$ as a function of n_s, and (d) $P(m_t | D, I)$ as a function of m_t.

where

$$P(q|I) = \frac{1}{\sqrt{2\pi}\sigma_q}e^{-q^2/2\sigma_q^2}, \tag{6.28}$$

N is a normalization constant that can be calculated using the definition of a probability density function, and σ_q is the estimated uncertainty on the jet energy scale. While trivial in principle, this integral is nontrivial in practice because one has to rebuild f_s and f_b at every step of the numerical integration over q. We generated two sets of 100,000 different values of q's (q_1 and q_2, to assign jet energy scales to f_s and f_b in a random manner), assuming σ_q to be 7.5 %, and restricting fluctuation in q to within $\pm$ one σ_q of q=0. In each step of the integration over q, $f_s(m_t, m_i^{fit}(q_1))$ and $f_b(m_i^{fit}(q_2))$ were recalculated by repeating the fitting procedure explained in the Section 6.2.2. (For the background, rescaled histograms are used as f_b.) We scanned the results of the fits, and confirmed that χ^2 was always less than 1.2, demonstrating the stability of results of the fit. Results for q-dependent f_s were also checked, and no evidence found for problems with the procedure.

The a posteriori probability for our data $P(m_t, n_s|D, I)$ is shown in Figs. 6.24 (a) and (b). There is no significant change from introducing the q-dependent a priori probability. Figure 6.24 (c) shows the a posteriori probability $P(n_s|D, I)$ as a function of the expected signal, n_s. The mean estimate of n_s and its RMS uncertainty becomes 19.6 $\pm$ 11.0 which is very similar to the result with q-independent prior. The a posteriori probability, $P(m_t|D, I)$ is shown in Fig. 6.24 (d) as a solid histogram. The dashed histogram is the same plot, but for the case when we do not consider the uncertainty in the jet energy scale (equivalent to $P(q|I) = \delta(q - 0)$), namely Fig. 6.23 (d). The mean value of m_t from a Gaussian fit to the peak (148.2 GeV/c^2) is shifted down by 1 GeV/c^2, and the standard deviation for the Gaussian fit is increased from 37.3 to 43.3 GeV/c^2 for the q-dependent formulation.
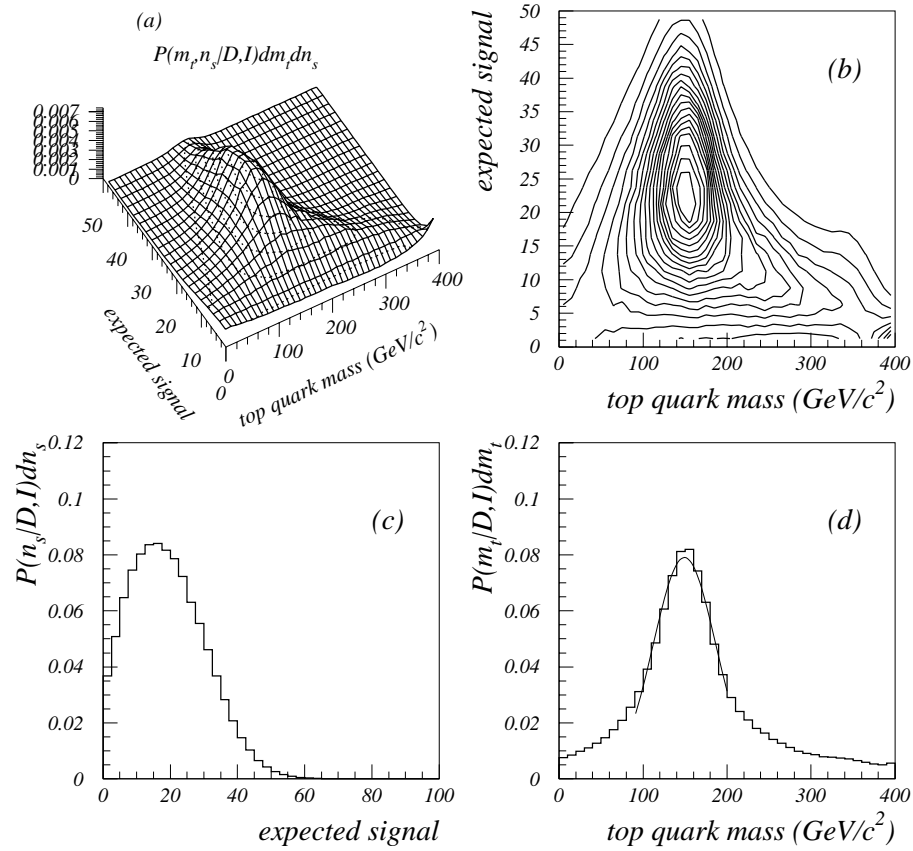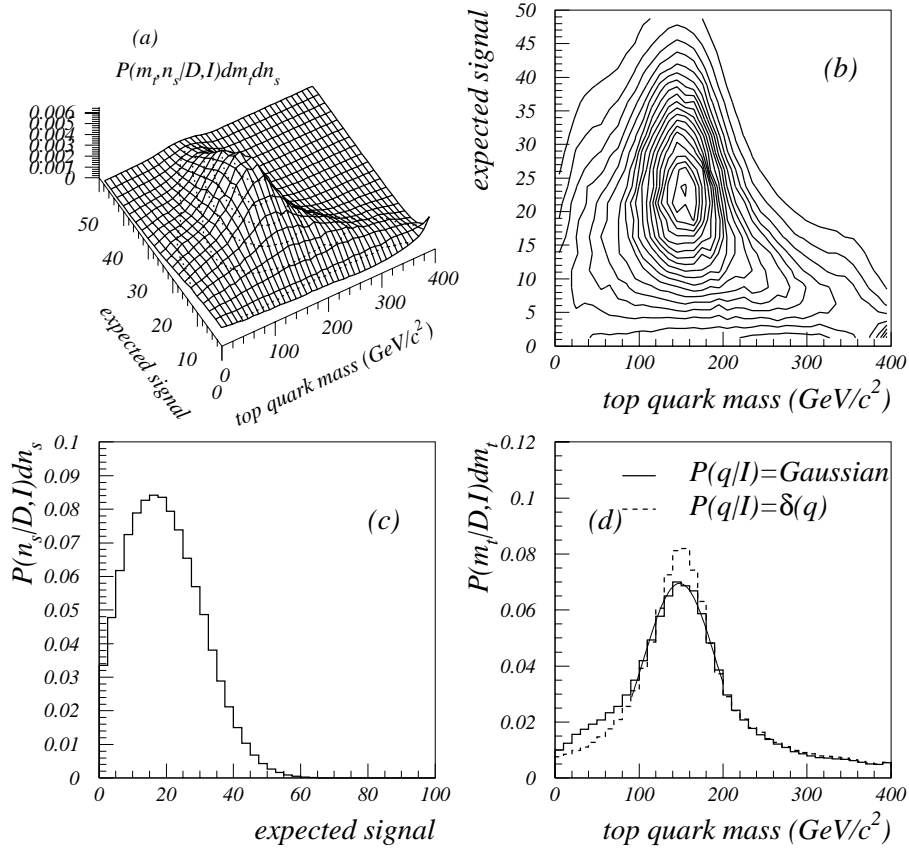
Figure 6.24: A posteriori probability distributions using an unbinned likelihood, a uniform prior for m_t, a Gaussian prior for n_b, and a Gaussian prior on the jet-energy scale. The standard 83 events from RUN I data were used in the analysis. (a) $P(m_t, n_s|D, I)$ as a function of m_t and n_s. (b) Contours of equal probability in (a), (c) $P(n_s|D, I)$ as a function of n_s, and (d) $P(m_t|D, I)$ as a function of m_t and solid line shows $P(m_t|D, I)$ as a function of m_t for the analysis including the dependence on scale (q), and the dashed line corresponds to the result of Fig. 6.23 (d).

Chapter 7

Conclusions

We performed a measurement of the production cross section of the top quark and of the mass of the top quark in the multi-jet final states. As described in Chapter 5, we observed a consistent excess of $t\bar{t}$ signal events using three different analyses, and found a cross section consistent with measurements reported by CDF [17] and DØ [18]. Nevertheless, the level of significance of the signal is not strong enough to establish the existence of the top quark in multi-jet final states. Based on the small excess of observed events above background, we pursued the measurement of the mass of the top quark. The lack of a significant excess of the signal events prevented us from estimating the mass of the top quark with precision.

We addressed our event classification problem by introducing three different techniques. The important-sampled grid search, the covariance matrix analysis, and the neural network analysis were discussed and compared. It was observed that, consistent with expectation, the neural network analysis provided the best performance. The $t\bar{t}$ production cross sections from individual analyses were consistent with measurements from other modes of $t\bar{t}$ production, but with rather

large uncertainties. We also addressed an alternative way of calculating the $t\bar{t}$ production cross section based on Bayesian probability theory. An upper limit (95 % confidence level) on $t\bar{t}$ production cross section at the top quark mass for 180 GeV/c² was estimated to be approximately < 30 pb for most sets of selection criteria.

Finally, we attempted to measure the mass of the top quark assuming that the small excess of events corresponded to $t\bar{t}$ production. An optimization procedure was developed in order to minimize the size of the statistical uncertainty of the measurement. However, due to a poor level of signal to background, as well as similarities in shape of background and signal in their fitted mass spectra, we were only able to estimate the mass of the top quark within relatively large uncertainties. We also addressed an alternative way of calculating mass of the top quark using Bayesian probability theory.

To summarize, out best estimate of the cross section for $t\bar{t}$ production, based on the all-jets channel, is 7.9 ± 7.1 pb (at $m_t = 160$ GeV/c²). Our best estimate of the mass of the top quark is 148 ± 52 GeV/c². These results are limited primarily by the statistics of the signal, following the analysis needed to increase the $t\bar{t}$ signal relative to the QCD background in the all-jets channel.

The upgrade of the DØ detector and of the Fermilab accelerator complex are in progress for the next collider run (RUN II, scheduled to begin 1999). We expect to have approximately 10 times higher luminosity than the luminosity in RUN I and to be able to identify b-quarks by observing secondary vertices in the upgraded SVX detector. With these improvements to the detector and the accelerator, we should be able to identify a few hundred $t\bar{t}$ events that produce multi-jet final states, and thereby test the predictions of the Standard Model for the branching ratio of the top quark into all-jets, and the mass of the top quark.

Bibliography

[1] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Benjamin/Cummings, (1983).

[2] F. Halzen and A. D. Martin, *Quarks and Leptons*, New York: John Wiley & Sons, (1984).

[3] J. F. Gunion, H. Haber, G. Kane, and S. Dawson, *The Higgs Hunter's Guide*, New York: John Wiley & Sons, (1990).

[4] S. Glashow, Nucl. Phys. **22**, 579 (1961).
S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
A. Salam and J.C. Ward, Phys. Lett. **13**, 168 (1964).

[5] V. D. Barger and R.J.N. Phillips, *Collider Physics*, Addison-Wesley Reading MA, (1987).
L. Ryder, *Quantum Field Theory*, Cambridge University Press (1987).

[6] M. Einhorn, *Current Physics - Sources and Comments*, **8** The Standard Model Higgs Boson (1991).

[7] UA1 Collaboration, Z. Phys. **C44**, 15 (1989).
UA2 Collaboration, Phys. Lett. **B186**, 440 (1987).

[8] CDF Collaboration, Phys. Rev. Lett. **75**, 11 (1995).

M. Demarteau, *QCD and High Energy Hadronic Interactions XXXIst Renoontres de Moriond*, Les Arcs, Savole, France (1996).

[9] SLC Collaboration, Phys. Rev. Lett. **63**, 2173 (1989).

[10] L3 Collaboration, Phys. Lett. **B231**, 509 (1989).
ALEPH Collaboration, Phys. Lett. **B231**, 519 (1989).
OPAL Collaboration, Phys. Lett. **B231**, 530 (1989).
DELPHI Collaboration, Phys. Lett. **B231**, 539 (1989).

[11] Particle Data Group, Phys. Rev. **D50**, (1994).

[12] P. Nason and S. Dawson, Nucl. Phys. **B303**, 607 (1988).

[13] E. Laenen, J. Smith, and W.L. van Neerven, Phys. Lett. **B321**, 254 (1994).

[14] E.L. Berger and H. Contopanagos, *The Perturbative Resummed Series for Top Quark Production in Hadron Reactions*, hep-ph/9603326.

[15] S. Catani, M. L. Mangano, P. Nason, and L. Trentadue, *The Top Cross Section in Hadronic Collisions*, CERN-TH/96-21, hep-ph/9602208.

[16] A.D. Martin, R.G. Roberts, and W.J. Stirling, Phys. Lett. **356B**, 89 (1995).

[17] CDF Collaboration, Phys. Rev. Lett. **74**, 2626 (1995).

[18] DØ Collaboration, Phys. Rev. Lett. **74**, 2632 (1995).

[19] J.M. Benlloch, N. Wainer, and W.T. Giele, *On the Search for the Top Quark in the Hardronic Decay Channel*, FERMILAB-PUB-93/060-T.

[20] F.A. Berends and H. Kuijf, Nucl. Phys. **B353**, 59 (1991).

[21] L. H. Orr , T. Stelzer , and W.J. Stirling, Phys. Rev. **D52**, 124 (1995). L. H. Orr , T. Stelzer , and W.J. Stirling, *Gluon radiation in $t\bar{t}$ production at the Tevatron $\bar{p}p$ collider*, hep-ph/9412294.

[22] B.R. Webber, *Hadronization*, Cavendish-HEP-94/17, hep-ph/9411394, lectures at Summer School on Hadronic Aspects of Collider Physics, Zouz, Suitzerland. August (1994).

[23] R.D. Field, *Applications of Perturbative QCD*, Addison-Wesley (1989).

[24] R.D. Field and R.P. Feynman, Nucl. Phys. **B136**, 1 (1978).

[25] F. Paige and S. Protopopescu, BNL Report No. BNL38034, (unpublished) release v. 6.49 (1986).

[26] G. Marchesini and B.R. Webber, Nucl. Phys. **B310**, 461 (1988).

[27] J. Thompson, *Introduction to Colliding Beams at Fermilab*, FERMILAB-TM-1909 (1994).

[28] C. W. Schmidt and C. D. Curtis, IEEE Trans. on Nuclear Science **NS-26**, 4120, (1979).

[29] R.J. Ducar, J.R. Lackey, and S.R. Tawzer, *FNAL Booster Intensity, Extraction, and Synchronization Control for Collider Operation*, FERMILAB-TM-1446 (1987).

[30] P.S. Martin, *Performance of the Fermilab Main Ring during the 1988-1989 Tevatron Collider Operation*, Proceedings of the 1989 IEEE Particle Accelerator Conference, 1829 (1989).

[31] Design Report, *The DØ Experiment at the Fermilab $\bar{p}p$ Collider*, (1984).

[32] S. S. Snyder, *Measurement of the Top Quark Mass at DØ*, Ph. D thesis, State University of New York at Stony Brook, Stony Brook, New York, (1995).

[33] DØ Collaboration, Nucl. Inst. and Meth. **A338**, 185 (1994).

[34] A.R. Clark et. al, Nucl. Inst. and Meth. **A279**, 243 (1989).

[35] F. Sauli, *Principles of operation of multiwire proportional and drift chambers* CERN Report 77-09, (1977).
In T. Ferbel, ed, *Experimental Techniques in High Energy Physics*, pages 79-188. Addison-Wesley, (1987).

[36] R.C. Fernow, *Introduction to experimental particle physics*, Cambridge University Press (1986).

[37] J. H. Cochran, *Search for the Truth in the eμ Channel at DØ*, Ph D thesis, University of New York at Stony Brook, Stony Brook, New York, December (1993).

[38] J.M. Butler, *Progress Report From DØ*, DØ Internal Note 1559, (1992).

[39] J.D. Jackson, *Classical Electrodynamics*, New York: John Wiley & Sons, (1975).

[40] L.T. Gross, *The Layperson's Guide to the TRD*, DØ Internal Note 2522, (1995).

[41] J.F Detoeuf, *et al.*, Nucl. Inst. and Meth. **A265**, 157 (1988).

[42] R. Wigmans, Annu. Rev. Nucl. Part. Sci. **41**, 41 (1991).

[43] C.W. Fabjan and R. Wigmans, *Energy Measurement of elementary particles*, CERN-EP/89-64 (1989).

[44] J. Womersely, *Everything your advisor should have told you about Calorimetry... But probably didn't know*, DØ Internal Note 2365 (1994).

[45] H. Aihara *et al.*, Nucl. Instr. and Meth. **A325**, 393 (1993).

[46] C. Brown, *et al.*, Nucl. Instr. and Meth. **A279**, 331 (1989).

[47] M. Abolins *et al.*, Nucl. Instr. and Meth. **A298**, 543 (1990).

[48] G.S. Gao and R. Partridge, IEEE Trans. Nucl. Sci. **NS-38**, 286 (1991).

[49] M. Fortner *et al.*, IEEE Trans. Nucl. Sci. **NS-38**, 480 (1991).

[50] M. Goosens *et al.*, CERN Library Program No. Q100 (1991).

[51] J. Butler, *Main Ring Deadtime*, DØ Internal Note 1682 (1993).

[52] N. Amos, *Main Ring Veto Counters for* RUN IB, DØ Internal Note 2072 (1994).

[53] M. Fischler, F. Rinaldo, and S. Wolbers, *Production Farms at Fermilab*, available from `http://fnhppc.fnal.gov/farms/info/`.

[54] D. Genser, *DPM Input Spooler for DØ data Processing on UNIX*, available from `http://d0sgi0.fnal.gov/d̃zero/software_adv.html`.

[55] M. Fausey *et al.*, *OCS Reference Guide*, Fermilab computing division product, GA0012.

[56] *RBIO Users' Guide*, Fermilab computing division product, SU0033.

[57] J. Shiers, *FATMEN - Distributed File and Tape Management*, CERN Program Library **Q123** (1994).

[58] Q. Li-Demarteau, *DØFIX FIX_DØRECO*, available from `http://d0sgi0.fnal.gov/physics_analysis/d0fix/d0fix.html`.

[59] DØ Collaboration, *DØ Software Documentation*, available from `http://d0sgi0.fnal.gov/software/software.html` (1995).

[60] UA1 Collaboration, Phys. Lett. **132B**, 214 (1983).

[61] S. Youssef, Comp. Phys. Comm. **45**, 423 (1987).

[62] A. Milder and R. Astur, *Jet Energy Calibration using the Missing E_T Projection Fraction*, DØ Internal Note 1657 (1993).

[63] T. M. Joffe-Minor and R. Astur, *A Study of the Effects of the CAFIX Energy Scale Corrections*, DØ Internal Note 2211 (1994).

[64] CDF Collaboration, Phys. Rev. Lett. **69**, 2896 (1992).

[65] M. Narain, *Electron identification in the DØ detector*, DØ Internal Note 1548 (1992).

[66] C. Gerber *et al.*, *Muon momentum determination*, DØ Internal Note 2140 (1994).

[67] C. Cretsinger, C. Stewart, and E. Won, *Search for the Top Quark in the All-jets Channel at DØ*, DØ Internal Note 2140 (1995).

[68] C. Cretsinger, *Search for the Top in the All-Jets Channel*, Ph D thesis, University of Rochester, Rochester, New York, December (1995).

[69] D. Edmunds, S. Gross, and P. Laurens, *Triggering on Large Tiles with the Level 1 Calorimeter Trigger*, DØ Internal Note 1680.

[70] C. Cretsinger and C. Stewart, *Preliminary Study of Parameters for Top to Jets Event Selection*, DØ Internal Note 2499 (1995).

[71] DØ Production Database Group, available from http://d0wop.fnal.gov/offline/pdb/get_filt_lum.html.

[72] DØ Collaboration, Phys. Rev. **D52**, 4877 (1995).

[73] H. Xu and R. Partridge, *Heavy-Quark Tagging at DØ*, DØ Internal Note 1997 (1993).

[74] J. Hobbs, *The DØ Run 1A $t\bar{t} \rightarrow \mu + jets + \mu(tag)$ Analysis*, DØ Internal Note 2389 (1994).

[75] C. Stewart, private communication (Now at Bio-logic Systems Corp. One Bio-logic Plaza Mundelein, IL 60060).

[76] T. Marshall and H. Haggerty, *Crud Removal from Muon Drift Chamber Wires*, DØ Internal Note 2556 (1995).

[77] T. Marshall, private communication.

[78] The DØ All-Jets Group, *Status Report: Search for the Top in the All-Jets Channel*, DØ Internal Note 2807 (1995).

[79] UA2 Collaboration, Z. Phys. **C36**, 175 (1987).

[80] A. Castro *et al.*, *Multi-jet Analysis for the Top Search at the Fermilab Collider*, FERMILAB-Conf-94/152-E (1994).

[81] F. V. Tkachov, *Calorimetric Measurement and Jet Counting*, FERMILAB-PUB-95/191-T (1995).

[82] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill (1969).

[83] H. Prosper, *et al.*, *The Random Grid Search: A Simple Way to Find Optimal Cuts*, At the 1995 Computing in High Energy Physics Conference, Rio de Janeiro, Brazil (1995).

[84] H. Prosper and R. Raja, *Analysis of $t\bar{t} \rightarrow e\mu$ events using the H-matrix*, DØ Internal Note 2035 (1994).
S. Beri, P. Bhat, R. Raja, and H. Prosper, *An H-matrix study of the $t\bar{t} \rightarrow e\mu$ channel*, DØ Internal Note 2227 (1994).

[85] N. Amos, private communication.

[86] C. Peterson and T. Rögnvaldsson, *An Introduction to Artificial Neural Networks*, CERN School of Computing, (1991).

[87] R. Beale and T. Jackson, *Neural Computing: An Introduction*, Adam Hilger, New York, (1991).
J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, (1990).

[88] C. Peterson and T. Rögnvaldsson, *JETNET 3.0 - A Versatile Artifical Neural Network Package*, CERN-TH.7135/94 (1994).

[89] D. Ruck *et al.*, IEEE Trans. Neural Networks **1** No. 4, 303 (1990).

[90] H. Prosper, *Some mathematical comments on Feed-Forward Neural Networks*, DØ Internal Note 1606 (1993).

[91] W. T. Eadie, D. Drijard, F. E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics*, American Elsevier Inc. NY (1971).

[92] E. K. Blum and L.K. Li, Neural Networks **4**, 511 (1991).

[93] W. Carvalho and T. Taylor, *Multiple Interaction Tool Study for Run I*, DØ Internal Note 2798 (1995).

[94] S. Ahn, private communication.

[95] E. Won, *On Matters of Significance*, DØ Internal Note 2465 (1995).

[96] R. T. Cox, Am. J. Phys. **14**, 1 (1946).

[97] E. T. Jaynes, *Probability Theory: The Logic of Science* (1996) (Unpublished), available from `http://www.math.albany.edu:8008/JaynesBook.html`.

[98] G. L. Bretthorst, *An Introduction to Model Selection Using Probability Theory as Logic*, in Maximum Entropy and Bayesian Methods, Santa Barbara, (1993).

[99] T. Nakamura and Y. Suto, *Probability Distribution of the Hubble Constant and the Age of the Universe Inferred from the Local Observations*, UTAP-202/95, also astro-ph/9505036 (1995).
Papers related with Bayesian statistics can be found from `http://xxx.lanl.gov/archive/bayes-an`.

[100] J. Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer-Verlag New York (1996).

[101] T. J. Loredo, *From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics*, Reprinted from: Maximum Entropy and Bayesian Methods, Kluwer Academic Publishers, Dordrecht, 81-142 (1990).

[102] G. D'Agostini, *Probability and Measurement Uncertainty in Physics - a Bayesian Primer*, DESY 95-242 (1995).

[103] G. D'Agostini, Nucl. Inst. and Meth. **A362**, 487 (1995).

[104] F. James, *The Statistics of Very Small Samples - Bayesian and Classical Approaches*, CERN Cours/Lecture Series (1986).

[105] M. A. Thomson, *The Use of Maximum Entropy in Electromagnetic Calorimeter Event Reconstruction*, CERN-PPE/96-055 (1996).

[106] M. Paterno, *et al.*, *A Recipe for the Construction of Confidence Limits*, DØ Internal Note 2775 (1995).

[107] CDF Collaboration, Phys. Rev. **D50**, 2966 (1994).

[108] A. G. Frodesen, O. Skjeggestad, and H. Tøfte, *Probability and Statistics in Particle Physics*, Columbia University Press, NY (1979).

[109] L. Lyons, *Statistics for Nuclear and Particle Physics*, Cambridge University Press, (1989).

[110] J. P. Berge, F. T. Solmitz, and H. D. Taft, Rev. Sci. Inst. **32**, 538 (1961). O. I. Dahl *et al.*, SQUAW Technical LBL Report P–126 (1968).

[111] E. Won, *A Preliminary Analysis of the Mass of the Top Quark using* SQUAW *in the All-jets Channel*, DØ Internal Note 2586 (1995).

[112] S. Chopra, R. Raja, *Estimation of the QCD background to* $W \to e\nu + jets$, DØ Internal Note 2098 (1994).

[113] F. Hsieh, private communication.

[114] T. Ferbel *et al.*, *Kinematic Fitting for the Mass of the Top Quark*, Proc. of the Workshop on Heavy Quarks, Rio de Janeiro, A. Santoro (ed), (1995).

[115] M. Pang, *Search for top in $\overline{p}p$ collisions at $\sqrt{}\ s\ =\ 1.8\ TeV\ by\ Constrained Kinematic Fitting*, Ph. D thesis, Iowa State University, Ames, Iowa, (1994).

[116] F. James, MINUIT *Minimization Package Reference Manual*, CERN Program Library Number D506 (1994).

[117] E. Won, *Update on the Extraction of the Mass of the Top Quark in Multijet Final States - Likelihood vs. Bayesian Estimators*, DØ Internal Note 2833 (1996).

[118] CDF Collaboration, Phys. Rev. Lett. **75**, 3997 (1995).

[119] C. Hill and S. Parke, Phys. Rev. **D49**, 4454 (1995).

[120] T. Ferbel and H. Prosper, private communication.

Appendix A

Fisher's Variable and Neural Networks

A.1 Effects on Kinematic Parameters of Selecting on Fisher's Variable and Network Output

In order to examine the impact of Fisher's variable defined in Eq. (5.1) and of network output on kinematic parameters, we compared the distributions of parameters before and after imposing these criteria.

Figures A.1 (a), (b), (c), and (d) show the distributions of $\mathcal{A}$, H_T^{3j}, $E_{T(5)}$, and $\mathcal{C}$, respectively, for $t\bar{t}$ ISAJET events with $m_t = 180$ GeV/c^2. The points have no selection on Fisher's variable, the light-shaded histograms require Fisher's variable to be greater than 0.3, and dark-shaded histograms require Fisher's variable to be greater than 0.6. Although cutting on Fisher's variable removes preferentially events that have small values of $\mathcal{A}$, the cut also removes also a substantial fraction

of events with higher values of $\mathcal{A}$. However, for distributions in H_T^{3j}, and $E_{T(5)}$, only events with low values are removed by selecting on Fisher's variable, as shown in Figs. A.1 (b) and (c). Similar behavior is observed in the distributions of $\mathcal{C}$.

Next, we performed the same exercise with the search sample, and the results are shown in Figs. A.2 (a), (b), (c), and (d). One distinctive feature with the search sample is that there are shoulders at the low values of H_T^{3j} and $E_{T(5)}$, as can be seen in Figs. A.2 (b) and (c). This may be the reason for the under-performance of the covariance matrix analysis relative to the grid analysis.

Similar studies were performed for selections on the output of the neural network. Figures A.3 and Figs. A.4 show distributions of kinematic parameters using $t\bar{t}$ ISAJET events of $m_t = 180$ GeV/c^2 and the search sample, respectively. Here, the selections on output at 0.7 and at 0.96 do not leave any excess of events at low values of H_T^{3j} or $E_{T(5)}$.

A.2 Details on Neural Networks

We rewrite the explicit analytic expression for F (Eq. (5.7)) for the network training described in Chapter 5:

$$F(x_1, ..., x_5) = g\left[\frac{1}{T} \sum_{j=1}^{8} \omega_{1,j} g\left(\frac{1}{T} \sum_{k=1}^{5} \omega_{j,k} x_k + \theta_j \right) + \theta_i \right] \tag{A.1}$$

where

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \equiv \begin{pmatrix} H_T \\ \mathcal{A} \\ \mathcal{C} \\ H_T^{3j} \\ E_{T(5)} \end{pmatrix}.$$
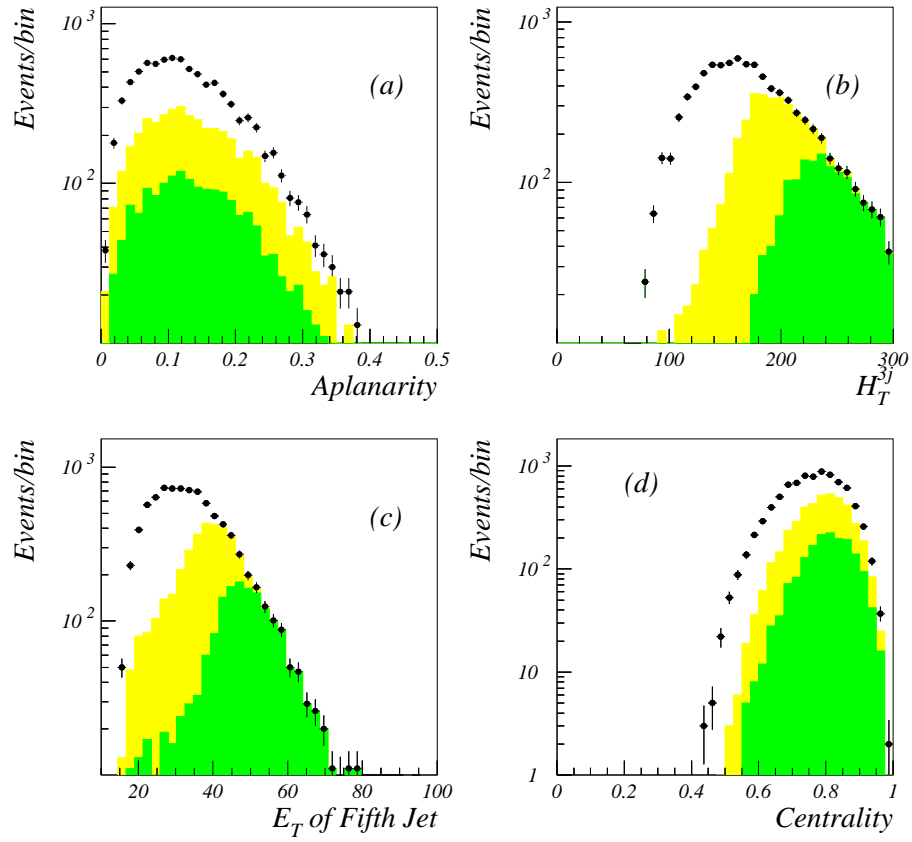
Figure A.1: Effect of Fisher's variable on the signal sample. (ISAJET of $m_t = 180$ GeV/c^2) Points show distributions with no selection on Fisher's variable, shaded histograms are Fisher's variable greater than 0.3, and dark-shaded histograms Fisher's variable greater than 0.6.
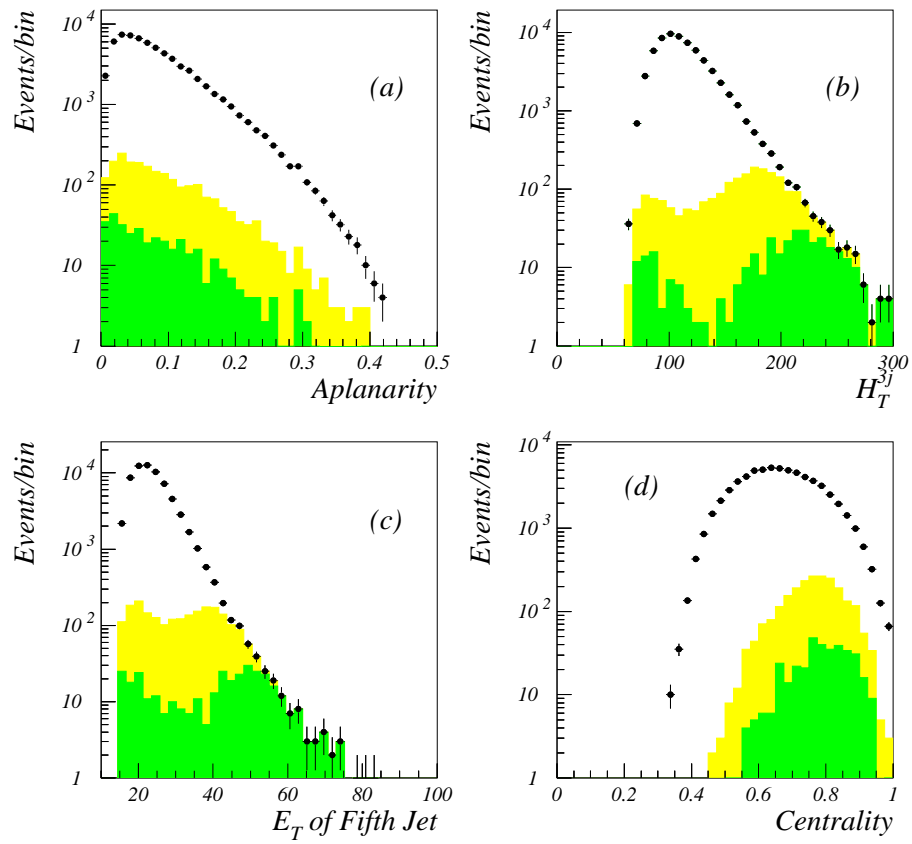
Figure A.2: Effect of Fisher's variable on the search sample. Points show distributions with no selection on Fisher's variable, shaded histograms are Fisher's variable greater than 0.3, and dark-shaded histograms Fisher's variable greater than 0.6.
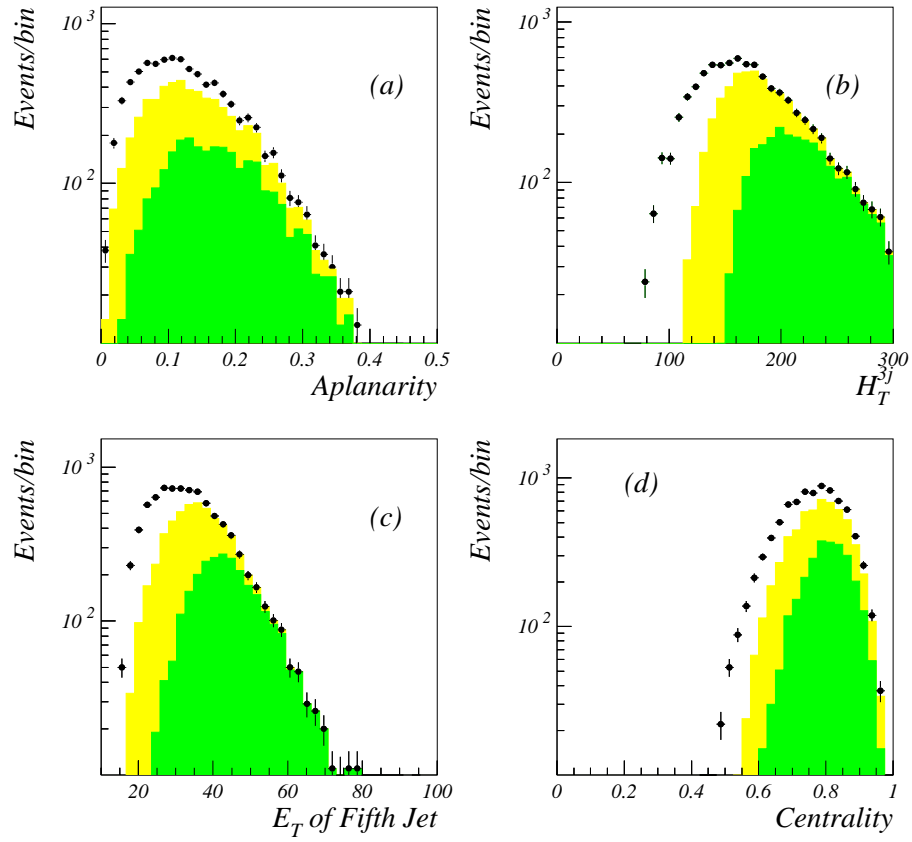
Figure A.3: Effect of the network output on the signal sample. (ISAJET of $m_t = 180$ GeV/c^2) Points show distributions with no selection on network output, shaded histograms are network output greater than 0.7, and dark-shaded histograms network output greater than 0.96.
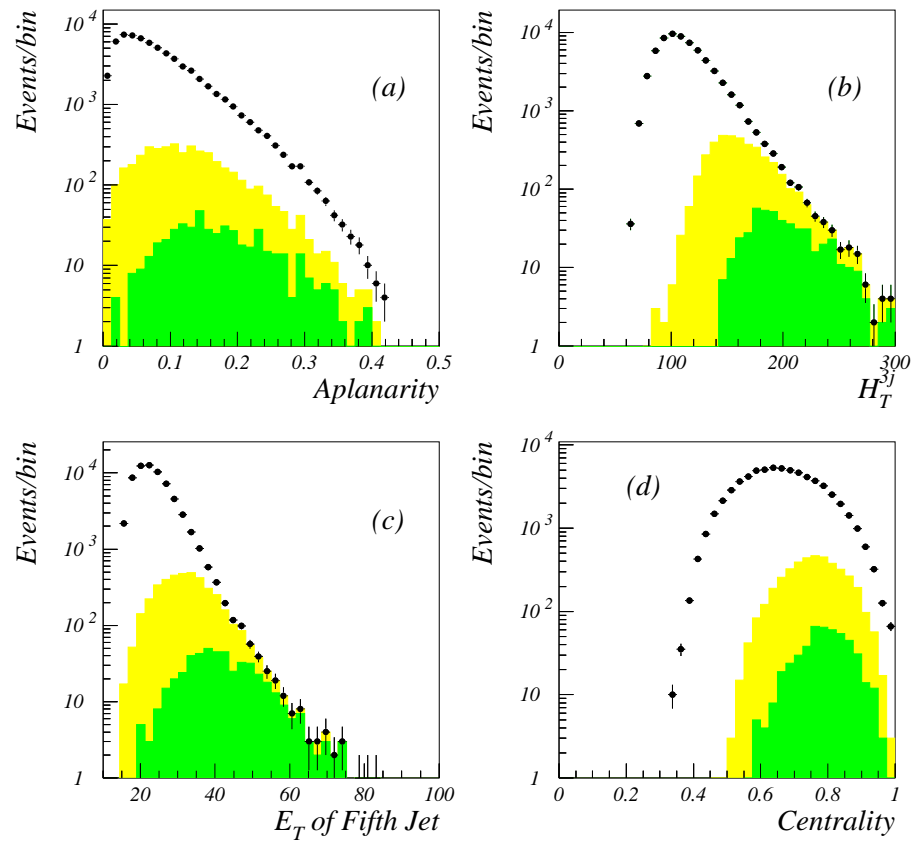
Figure A.4: Effect of the network output on the search sample. Points show distributions with no selection on network output, shaded histograms are network output greater than 0.7, and dark-shaded histograms network output greater than 0.96.

The numerical results of the training, given in Chapter 5, were:

$$\omega_{j,k} = \begin{pmatrix} +0.207 & +0.342 & -0.025 & -0.263 & +0.193 \\ -2.992 & +0.514 & +2.036 & -0.598 & -0.875 \\ +2.917 & -5.573 & -2.074 & -2.871 & -0.827 \\ -5.354 & +0.562 & +0.487 & +3.263 & -0.472 \\ +0.198 & +0.992 & +0.714 & -0.082 & +0.673 \\ -0.294 & +2.348 & +2.609 & -0.450 & +0.496 \\ +1.602 & +0.900 & +1.316 & +0.014 & +2.312 \\ -1.169 & -0.247 & +0.743 & -1.097 & -0.447 \end{pmatrix}$$

$$\omega_{1,j} = \begin{pmatrix} +0.204 & -1.414 & -1.711 & -2.533 \\ +0.391 & +1.801 & +1.926 & -0.771 \end{pmatrix}$$

$$\theta_j = \begin{pmatrix} -1.001 & +1.313 & +1.171 & +1.346 \\ -1.363 & -0.716 & -3.835 & +0.932 \end{pmatrix}$$

and

$$\theta_1 = -0.577.$$

To visualize the variations in the size of weights, we sketched the 5-8-1 architecture, with lines whose widths are proportional to the absolute magnitude of their weights. Figure A.5 shows the results. The sizes of the gray circles for the hidden nodes are proportional to the absolute magnitude of threshold values, θ_j. It is clear that H_T, $\mathcal{A}$, $\mathcal{C}$, and $E_{T(5)}$ appear to provide greater weights to the hidden nodes that have greatest impact on the network output.
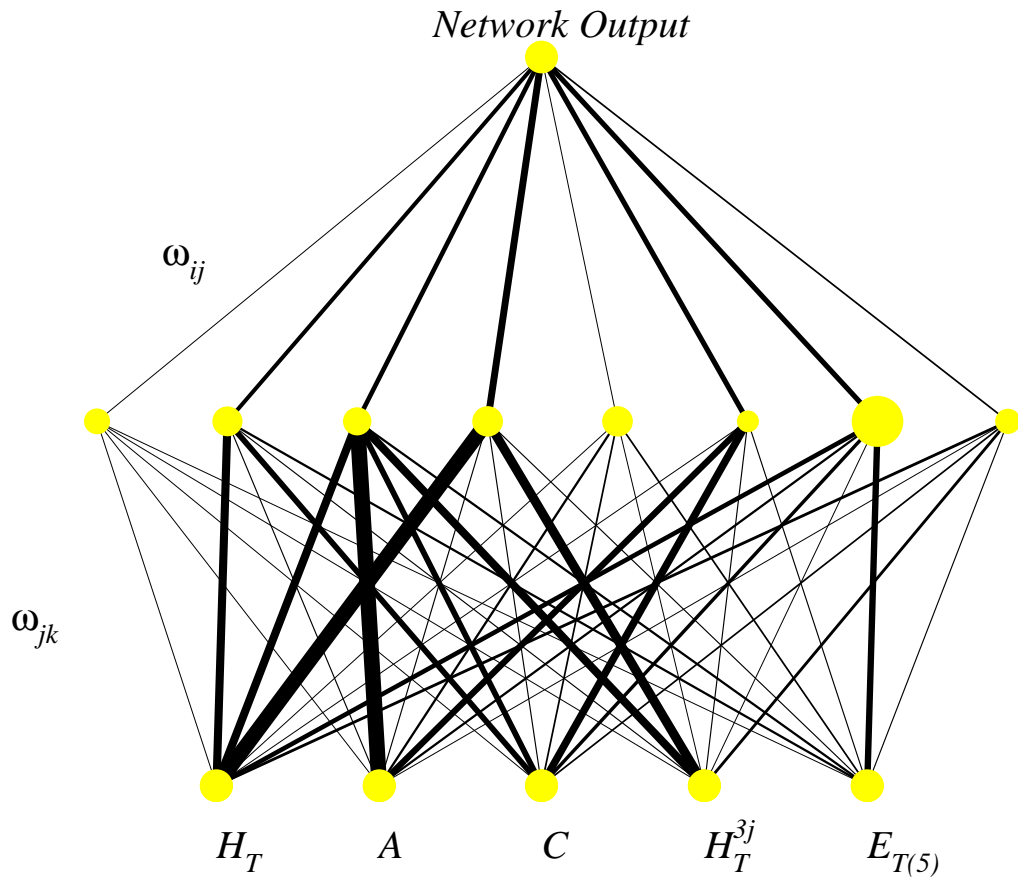
Figure A.5: Relative size of weights among networks. The absolute values of weights are proportional to width of lines between nodes, and the sizes of circles for the hidden nodes are proportional to the absolute magnitude of threshold values θ_j.

Appendix B

Back-Propagation Learning

As was pointed out in Chapter 5, the minimization of the mean square error function E is achieved by updating of weights and thresholds using the method of gradient descent[86]:

$$\Delta \omega = -\eta \frac{\partial E}{\partial \omega} \tag{B.1}$$

here η is the learning rate parameter $(0 < \eta < 1)$ and the ω is the parameter contains implicitly all the weights and thresholds. Because η is a positive quantity, the gradient descent method finds a minimum value of E.

In what follows, we discuss how the back-propagation is accomplished. In order to keep the calculation simple, we assume that the weights $\omega_{j,k}$ include threshold values, $T = 1$, and we suppress the $1/N_p \sum_p$ in Eq. (5.9). Also, we replace F_i with o_i in order to emphasize that the network is not yet trained. Therefore, the mean square error function in Eq. (5.9) and the network output o_i can be rewritten as:

$$
\begin{aligned}
E &= \frac{1}{2} \sum_i (o_i - t_i)^2, \\
o_i &= g \left[\sum_j \omega_{i,j} g \left(\sum_k \omega_{j,k} x_k \right) \right].
\end{aligned}
\tag{B.2}
$$

With the chain-rule of differentiation, one gets for the weights "connecting" output-to-hidden nodes:

$$
\begin{aligned}
\Delta\omega_{i,j} &= -\eta\frac{\partial E}{\partial\omega_{i,j}} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{(B.3)}\\
&= -\eta\sum_l(o_l - t_l)\frac{\partial o_l}{\partial\omega_{i,j}}\\
&= \eta\sum_l(t_l - o_l)\frac{\partial}{\partial\omega_{i,j}}g\left[\sum_m\omega_{l,m}g(m)\right]\\
&= \eta\sum_l(t_l - o_l)g'(l)\sum_m\delta_{il}\delta_{jm}g(m)\\
&= \eta(t_i - o_i)g'(i)g(j)\\
&\equiv \eta\delta_i g(j)
\end{aligned}
$$

where $\delta_i \equiv (t_i - o_i)g'(i)$, the i in $g(i)$ denotes the relevant index in the function g (as opposed to the dummy index), and the symbol δ_{il} denotes the Kronecker delta ($\delta_{il} = 1$ if $i=j$, otherwise $\delta_{il} = 0$). For the weights connecting hidden-to-input nodes, the rule of gradient descent gives:

$$
\begin{aligned}
\Delta\omega_{j,k} &= -\eta\frac{\partial E}{\partial\omega_{j,k}} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{(B.4)}\\
&= \eta\sum_i(t_i - o_i)g'(i)\frac{\partial}{\partial\omega_{j,k}}\left[\sum_l\omega_{i,l}g(\sum_m\omega_{l,m}x_m)\right]\\
&= \eta\sum_i\delta_i\sum_l\omega_{i,l}\frac{\partial}{\partial\omega_{j,k}}\left[g(\sum_m\omega_{l,m}x_m)\right].
\end{aligned}
$$

One should note that $\omega_{i,l}$ in the above equations are the weights and thresholds connecting output to hidden nodes, and $\omega_{j,k}$ are hidden to input nodes. That was used in the last step in Eq. (B.4). Now, using the fact that

$$
\frac{\partial}{\partial\omega_{j,k}}\left[g(\sum_m\omega_{l,m}x_m)\right] = g'(l)\sum_m\frac{\partial\omega_{l,m}}{\partial\omega_{j,k}}x_m \quad\quad\quad\quad \text{(B.5)}
$$

$$= g'(l) \sum_m \delta_{jl} \delta_{km} x_m,$$

one finally obtains:

$$\Delta \omega_{j,k} = \eta \sum_i \delta_i \sum_l \omega_{i,l} g'(l) \sum_m \delta_{jl} \delta_{km} x_m \qquad \text{(B.6)}$$
$$= \sum_i (t_i - o_i) g'(i) \omega_{i,j} g'(j) x_k.$$

Therefore, updating via gradient descent becomes:

$$\Delta \omega_{j,k} = \eta \sum_i \delta_i \omega_{i,j} g'(j) x_k \qquad \text{(B.7)}$$
$$\equiv \eta \delta_j x_k$$

where $\delta_j \equiv g'(j) \sum_i \omega_{i,j} \delta_i$. The information at the output to hidden layers (the term δ_i) is consequently "back-propagated" through the network (δ_j)[87]. Using Eq. (B.3) and (B.7), one can implement the back-propagation algorithm[88].